The Art of Counting

Enumeration and combinatorics

Everything is mathematical

The Art of Counting

The Art of Counting

Enumeration and combinatorics

Juanjo Rué

Everything is mathematical

To my parents
and to my grandmother Antonia.

Published by RBA Coleccionables, S.A.
c/o Hothouse Developments Ltd
91 Brick Lane, London, E1 6QL

Localisation: Windmill Books Ltd.

ISSN: 2050-649X

Printed in Spain

Contents

Preface

One sunny afternoon I was walking through the Luxembourg Gardens in Paris with a female friend. It was all very busy: highly skilled pétanque players, children planning future misdeeds and couples displaying their affection before all and sundry. The mere fact that we were in the Latin Quarter of Paris with all its history gave me a feeling of jubilation. A place like that provides inspiration for reflection and debate. And so it was that we began to talk. But we did not understand each other: I spoke about mathematics, or rather, I spoke in mathematical language, although not about the concepts that were behind those technical terms. I understood that there was a problem here, and I decided to find out what that problem was.

Mathematics involves techniques (a lot!) but it is basically about ideas. And the ideas must be expressed in a simple way, even though that process of simplifying can become an arduous exercise. It is in this issue of simplification that those of us who work in this field often give out the wrong message, creating an image that is far from reality. In my view, the experienced researcher (and the not-so-experienced one too) has a commitment, or rather, a moral duty, to society to feed back the essence of their work – the good ideas and the beautiful results. That is my modest goal in this modest book.

This book is dedicated to combinatorics, a discipline for which one only has to know how to draw and count to be able to deduce very ingenious and unexpected results. It is a discipline containing problems with simple enough formalisations – easy enough even for children to understand – but which can be impossible to solve. And in all cases, it has problems and notions that help give us a better understanding of reality.

To explore this area we shall begin in a simple fashion – by counting on our fingers. These calculations will allow us to understand why in some situations it is best to bet on games of chance, even though we might know that the odds are against us. Next we shall go on to see how combinatorics goes further than just counting and how it also covers the study of discrete models which simplify reality. At the same time we shall look at how, when a road network is being planned, the building of the bridges will depend on how the towns are to be linked, and not on the skill of the chief engineer.

All the notions we shall introduce will be handled in a very special way by a character who was to mark a turning-point in what today we understand as

combinatorics, the brilliant and prolific Paul Erdös. He will be the one to guide us through present-day combinatorics. That includes counting without using our fingers, exploring chaotic objects in which we will be able to find hidden structures, and exploring where chance and determinism meet – the combinatorics of playing dice with graphs and meeting up with old friends at school reunions. God not only plays dice with the Universe; He does with many other things as well.

Finally we shall show that in the arithmetic of numbers there is also an underlying combinatorics, which could almost be described as magical, plus, we shall explain one of the most important mathematical theorems of recent years: the Green–Tao theorem, to some the greatest mathematics result of all time.

I would also like to take this opportunity to thank Guillermo Navarro for all the suggestions he gave me through this long writing process, and Javier Fresán for trusting in my skills as a writer. I would also like to thank Marta Benages and Jose Miguel No for reading part of the preliminary drafts for this book.

Chapter 1

Let's Count!

When we journey to other lands we should make an effort to try out the best cuisine on offer in the different regions we visit. Travel, in the true meaning of the word, must include making culinary discoveries. That is surely so, as a region's typical dishes allow the palate to share and enjoy the traveller's adventures. While enjoying the traveller's rest, experimenting with the local delicacies is an essential part of understanding the culture and customs of the places being explored.

Unfortunately, it is often the case that the palate is too accustomed to the food of its homeland and is unable to face up to the gastronomic challenges it encounters. As a result, a temporary truce is declared and the traveller rather guiltily enters one of the many Italian restaurants dotted around the globe. And in all Italian restaurants worthy of the name, the star dish is pizza. On reaching this point, dilemmas fundamental to human existence make their appearance: shall we have four cheeses pizza or four seasons pizza? Shall we try a medium-sized one, or just a small one? Shall we order it with or without onions? Shall we add some extra toppings? Faced with such a variety, hesitation is inevitable because far beyond the pizzas as offered on the menu, a whole universe of combinations opens up before the traveller.

With such diversity, it is natural to wonder just how many different pizzas could be ordered within the resources that the restaurant has at its disposal. The count is complicated by several factors. For example, there are always incompatible toppings, those combinations that could never work when put on the same pizza.

The art of counting

But the important issue here is that similar questions arise in daily life. How and in how many ways can you combine ties from your wardrobe so that your outfit is never repeated and, more importantly, for there to be a certain degree of matching of colours. Or, for example, how many different routes could we take for a drive in the

country near to your home. These questions form part, to a greater or lesser degree, of what is known as *enumerative combinatorics*. This discipline is the art of counting, and counting is an art that requires its own techniques.

The art of counting is not a new discipline; on the contrary, it is one of the most natural and fundamental aspects of mathematics. Just as in many other areas of science, its origin is not clear. In the West, a systematic study of counting was started as a consequence of its close relationship with the calculation of probabilities. Back in the 17th century, Blaise Pascal (1623-1662) and Pierre de Fermat (1601-1665) pondered over games of chance and questions of probability that had been raised by a gambler desperate for answers – Chevalier de Méré. In these questions, the calculation of the most favourable cases divided by the possible cases (what would subsequently become known as *Laplace's rule of succession*) was the key element for giving precise explanations for the observed phenomenology. Later, great scientists such as Gottfried Wilhelm Leibniz (1646-1716) and Jakob Bernoulli (1654-1705) managed to establish combinatorics as a discipline independent from other branches of mathematics. At a later date, with the evolution of mathematical language and its techniques, it was noted that combinatorics were closely linked to another type of problem, apart from those of probability and counting: Leonhard Euler (1707-1783) was the first to come up with the notion of using graphs when he proposed using them to study the problem of the bridges of Königsberg, and Arthur Cayley (1821-1895), while researching isomerism in saturated hydrocarbons, obtained numerous results in the enumeration of graphs. Combinatorics is nowadays an area of intense scientific activity, both at a theoretical level and in practical areas.

Far beyond just enabling us to count on our fingers (on many occasions, with great skill), enumerative combinatorics allows us to access far more qualitative and structural information on the objects we are studying. That is the reason why, in spite of the fact that enumerative combinatorics is a discipline in its own right in the Mathematics Olympiad, many other fields use its techniques to arrive at conclusions. In probability, knowing how to count is fundamental to working out the proportion of favourable cases compared to the possible cases. And in building algorithms, it is of fundamental importance to know how many steps a certain process needs so as to decide if the new one is better or worse than already existing methods. Far beyond these disciplines, knowing how to count has profound ramifications in more heterogeneous areas such as mechanical statistics and enumerative geometry.

Jakob Bernoulli (above on the left), Leonhard Euler (left) and Gottfried Wilhelm Leibniz, forerunners of combinatorics, immortalised on these postage stamps.

RAYMOND LULL, ANOTHER PIONEER OF COMBINATORICS

Raymond Lull (1232–1315) was born into a wealthy Majorcan family shortly after James I of Aragon conquered the island. Throughout his life he showed great dedication to numerous varied activities. In his youth he was a member of James I's court and lived in circles that could be described as somewhat hedonistic. Some years later, after experiencing a number of visions of Jesus Christ on the cross, he abandoned his family and devoted himself to contemplation. He became a missionary, a Franciscan, and an ascetic, and travelled throughout the Mediterranean preaching Christianity and conversion. His work spans both philosophy and theology, taking in astronomy and alchemy. He is in fact considered to be the father of many present-day disciplines such as combinatorics. In his greatest work, *Ars Magna* (1305), Lull designs a method of combining religious and philosophical attributes, each of which can be selected from a list. To a certain extent, he got this idea of *combining* from the tools the Moors used in astronomy and navigation, which were based on a system of combining different positions with the aim of reaching their final conclusion. With this method, Lull aimed to discover the appropriate concepts for making judgements and syllogisms in any field, and to construct logical reasoning by using mathematical principles. His goal could perhaps be described as that of mechanising and *mathematising* knowledge from a premise based on combinatorics. Very much from the perspective of combinatorics, Lull is also considered to be one of the forerunners of computation and artificial intelligence.

Counting on your fingers

To start us off on our stroll around the world of combinatorics we shall use our fingers and one or two other common sense arguments. We might, however, need n fingers to reach our goal! As we shall see, there are objects that can be counted directly, others which we can count indirectly, and there will even be some that we shall be able to count in different ways. In each case, we shall show that the arguments used are not lacking in ingenuity.

The first step to getting a precise understanding of the basic principles of combinatorics is to know how to translate everyday language into generic and universal statements. We shall start off this process by translating the concepts 'and' and 'or' into the mathematical world. Let's look at a simple example so as to clarify this point. Suppose we go to our favourite restaurant for lunch. As we had breakfast very early in the day, our stomach is calling out for a culinary feast, both for a starter and a main course. However, if, due to the excesses of previous months, our dear GP has put us on a strict diet, then we might choose to have just one dish, either a starter or a main dish.

Let's look at the first case. In this case, the 'and' means that we must choose two dishes, one for the starter, and another for the main course. To make this choice we can follow this procedure: first we choose the starter that we most fancy, and once that has been done we choose the main dish we most fancy. This is simply a pair (starter, main course) that uniquely defines our lunch.

The abstract mathematical construction associated with this situation is as follows. Given two groups A and B, we define the concept of 'Cartesian product of A and of B' as a new set C, which is written as $A \times B$, the elements of which are *pairs*, where the first component of the pair belongs to A, while the second one belongs to B. In the case of our example, set A is the one for starters, B is the one for the main course and C is the total number of possible combinations for starters and main courses that we could choose. Each of these combinations of starters and main course are, therefore, pairs, (a, b) where a is an element of A, and b is an element of B. How many of these pairs exist or, what comes to the same thing, how is the number of elements of C calculated? The answer is obvious – by multiplying the number of elements contained in A by the number of elements contained in B. In the case of the example above, for each different choice of starter, I can choose any of the main dishes. This argument is completely general and covers what in combinatorics is known as the *multiplication principle*:

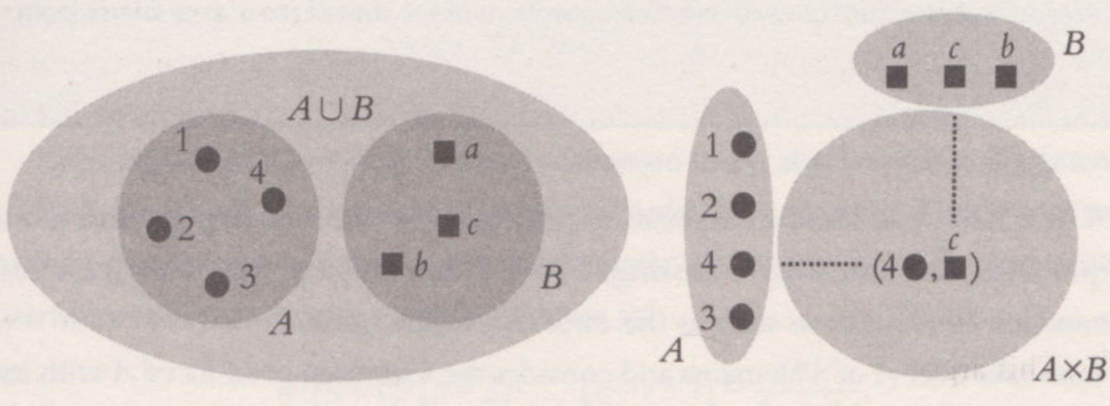
"Let A and B be two finite sets. Then the number of elements of the Cartesian product of A and B is equal to the number of elements of A multiplied by the number of elements of B."

Let's now take a look at the second situation, in which our spartan diet only allows us one dish. In how many ways can we make this choice? The number of possibilities will be precisely the total number of starters offered plus the number of main courses, provided there is no starter available that is at the same time a main course. This fact is translated mathematically in the following way. We consider two sets that are *disjunctive* A and B; in our case they are the starters and the main courses that the restaurant is offering, because in our hypothesis there is no dish that is both starter and main course. We want to choose an element from one of the sets, in other words, we want to choose an element from A or from B. This argument is the same as joining the sets A and B and choosing an element from the union.

What this means is that the colloquial language 'or' is translated into a disjunctive union in mathematical language. All this argument can be summed up in the combinatorics principal known as the *addition principle*:

"Let A and B be two finite sets, with no element in common. Then the number of elements of the union of A and B is equal to the sum of the number of elements of A and the number of elements of B."

The following diagrams are a graphic demonstration of the meaning of the operations of union and those of Cartesian products in sets, and the two principles that we have introduced here are deduced from them.



MATHEMATICAL NOTATION FOR DESCRIBING THE BASIC PRINCIPLES AND THE PRINCIPLE OF INCLUSION–EXCLUSION

The multiplication and the addition principles can be translated by using mathematical notation. By using $|A|$ to denote the number of elements of A, the multiplication principle is written in the following way:

$$|A \times B| = |A| \cdot |B|.$$

By bearing in mind that the union of A and B is written as $A \cup B$, the addition principle tells us that if A and B are disjunctive, then it holds that:

$$|A \cup B| = |A| + |B|.$$

The addition principle can be refined so as to obtain more information when the sets A and B have some common element. Let's suppose that A and B are not disjunctive, and therefore the intersection of A and B contain some element (mathematically this fact is denoted by writing that $A \cap B \neq \varnothing$). By bearing in mind that by adding the cardinal of A and that of B we are adding twice the number of elements common to the two sets, we can easily deduce the formula:

$$|A \cup B| = |A| + |B| - |A \cap B|.$$

By using the same arguments, this formula is generalised for three sets: let A, B, C be finite sets. Then the following relationship between the cardinals holds:

$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|.$$

This argument can be generalised to arbitrary families of sets to obtain what is known in combinatorics as the principle of inclusion-exclusion. This principle has numerous applications in combinatorics and number theory when, for instance, we want to estimate the size of the union of a series of sets of which we have estimations of the sizes of the mutual intersections.

The addition and the multiplication principles are the first step towards getting deeper into the world of counting. These elementary principles give rise to elementary constructions such as the ones that follow now.

Let's take a set A of n elements and consider the Cartesian product of A with itself a certain number of times (let's say r times). This new set if formed by sequences of

the type $(a_1, a_2,..., a_r)$ – in the mathematical jargon they are called *r-tuples* or *vectors* of length r) – where each of the components is an element of A. So, how many r-tuples are there? If r were equal to 2 we would be right inside the case of the multiplication principle, giving rise to $n \cdot n = n^2$. If r is different from 2, just by successively applying the multiplication principle it can be deduced that the number of r-tuples is equal to $n \cdot n \cdot ... \cdot n = n^r$. This construction is what is known as *variations with repetition*.

Now suppose that we complicate the problem a little more, and that we want to count r-tuples, but with the condition that all the components are different. The first important observation for solving this question is that the value of r, the length of the sequence, must be less than or equal to n, the number of elements of A, as otherwise there would be repetition of coordinates. To obtain this count we should observe that any vector $(a_1, a_2,..., a_r)$ with different components can be built in the following way: from among all the possibilities we choose the element corresponding to the first component, a_1. We next choose the second element from among all the possible ones, but without being able to choose a_1, as we have already chosen that for the first option. For the third component we can apply the same argument, ruling out the choice of a_1 and of a_2. And so on up to r. So what is the count? Well, the choice of a_1 is free, so we have n possibilities. For a_2 there is not so much freedom, as this one cannot be the element chosen for the first component. We therefore have $n-1$ possibilities for a_2. And so on. If we work it out, we get what are called *variations without repetition of length r*, whose formula is:

$$n \cdot (n-1) \cdot (n-2) \cdot ... \cdot (n-r+1).$$

In the particular case of making the value of r equal to the value of n, we get what is known as *permutation of n elements*. In this case, the formula for variations without repetition is written as:

$$n \cdot (n-1) \cdot (n-2) \cdot ... \cdot 2 \cdot 1.$$

This expression is normally written shortened to $n!$, also known as the *factorial of n*. The factorial tells us the number of ways of ordering a given set. This is very useful formula that is very widely used in mathematics. By using factorials, for example, variations without repetition can be written as:

$$n \cdot (n-1) \cdot (n-2) \cdot ... \cdot (n-r+1) = \frac{n!}{(n-r)!}.$$

The factorial can be defined uniquely for positive natural numbers, and for 0 the convention of defining $0! = 1$ is used. (There is an underlying reason for taking this definition, related to some formulae that are more general than the factorial and which stem from what is called *gamma function*, which is an essential element in diverse fields of mathematics, such as probability and the analytic number theory.)

The natural step to take now is to continue with disordered structures. Up to now we have been working with counting where the order matters, as all the time we were considering ordered r-tuples of elements in a given set A. However, there is an infinity of examples in which the order does not matter to us. Suppose that we want to go hiking in the mountains, and to do so we need to take the car (a 5-seater), but there are ten of us wanting to go on the trip. What matters is to find out which persons are chosen, and not the order in which they are chosen. How can we count disordered objects from ordered ones? In actual fact, there will not be much of a problem – a disordered object is an ordered object for which we have forgotten the order.

Let's inject some rigour into those last comments. A subset of r elements of A is a grouping of r elements (of the ns that A has), in which the order is of no importance. Given a set A with n elements, we want to know how many subsets of A there are that have r elements. This number is known as the number of *combinations without repetition* of size r in a set of size n. So, for each subset of size r we can define a total of $r!$ different r-tuples, where the $r!$ is obtained as a result of permuting the order of the elements (remember the permutations that we have just defined). As we have previously obtained the number of variations without repetitions, we can deduce that the number of combinations without repetition is equal to the number of variations without repetition of r elements, divided by the number of permutations. In other words, we obtain the following formula for the number of combinations:

$$\frac{n!}{r!(n-r)!}.$$

Mathematics has its own name and notation for this formula. That is where what is known as the *binomial coefficient* comes in, which is denoted as follows:

$$\binom{n}{r}.$$

These binomial numbers also count what we call *permutations with repetition*. Let's take a set A with two elements. For the sake of convenience let's say that they are

0 and 1. What is, then, the number of ordered sequences of length n with exactly r zeros and $n-r$ 1's? For example, if we take $n=4$ and $r=2$, we have the following six 4-tuples:

$$(0,0,1,1), (0,1,1,0), (0,1,0,1), (1,1,0,0), (1,0,0,1), (1,0,1,0).$$

Instead of directly counting how many permutations with repetition there are, what we will do is to take advantage of the fact that we know how to count the combinations without repetition and uniquely associate each permutation with repetition to a subset of $B=\{1,2,\ldots,n\}$ which is of size r. In this way we are perfecting matching up a permutation with repetition with a subset of r elements. Each sequence of length n defines a subset of $\{1,2,\ldots, n\}$ in the following way: if the figure in position m of the r-tuple is equal to 1, then m belongs to the subset; otherwise, it will not belong to the set. In the example that we have looked at we get the following correspondences:

$$\begin{gathered}(0,0,1,1)\to\{3,4\}\\(0,1,1,0)\to\{2,3\}\\(0,1,0,1)\to\{2,4\}\\(1,1,0,0)\to\{1,2\}\\(1,0,0,1)\to\{1,4\}\\(1,0,1,0)\to\{1,3\}\end{gathered}$$

The reader can easily see that this operation is invertible, and that for each subset of r elements we can construct a sequence with 0's and 1's which has exactly r 1's. This method that we have used here is called the *bijective method,* and later it will be very important for counting objects that are seemingly very different but which in reality are essentially the same.

A combination of the preceding ideas gives rise to the following result, known as the *committee problem*. Let's suppose that we want to form a committee of representatives in a group of n people, with the condition that the number of members in it can be arbitrary; the committee can even be left vacant. The question is to find how many ways we can make up that committee. To answer this question we are going to look at the problem in two different ways.

On the one hand we can reason that a certain committee will be made up of r persons, where r can vary between 0 (in the event that the committee has been left vacant) and n (in the case that all the group form part of the committee). With the value r fixed, how many committees of r members can be formed within a group of n persons? This is precisely the number of subsets of size r in an initial set of size n, or, as we said previously, that count is given by the binomial

$$\binom{n}{r}.$$

Since the number of committees is the total for all the choices of r, by the addition principle the result is that the number wanted is equal to the sum:

$$\binom{n}{0}+\binom{n}{1}+\ldots+\binom{n}{n-1}+\binom{n}{n}.$$

The dots in this formula tell us that the lower index of the binomials varies between 0 and n. Let's look at how we can obtain the same result in a different way. Let's begin by labelling the people, from 1 up to n. Each of them is identified by a different number. Now let's take a vector of length n in which each component is either 0 or 1. Each of these vectors gives us the following information: If, in a position i we wrote 1, then the person whose label is i will form part of the committee. And, inversely, if in position i we wrote 0, that person will not form part of the committee. For example, for $n=4$ one of these vectors would be (0,0,1,1), which tells us that the persons labelled 3 and 4 are the ones who will form the committee. We can now apply what we studied previously to find the number of vectors of length n with coefficients that are either 0 or 1. That is simply the number of variations with repetition of a set of size 2 (the 0 and the 1) as many times as n. That is, we find that the value is equal to 2^n.

The two values that we have obtained (the sum of binomials and the exponential) count the same number of ways in which we can form a committee in a population of n members. To sum up, we have proved the following combinatorics property:

$$\binom{n}{0}+\binom{n}{1}+\ldots+\binom{n}{n-1}+\binom{n}{n}=2^n.$$

THE BINOMIAL THEOREM AND ITS ENUMERATIVE CONSEQUENCES

A consequence of being able to count unordered objects is the algebraic result that follows, known as Newton's binomial in honour of the first person to use it systematically, the brilliant physicist and mathematician Isaac Newton (1642–1727). Let's take the binomial $1+x$. Then the polynomial $(1+x)^n$ is written as a polynomial in the form $a_0+a_1 x+a_2 x^2+\ldots+a_n x^n$. To calculate the coefficients we can apply a combinatorics argument. Let's write $(1+x)^n$ as n products of the binomial $1+x$:

$$(1+x)\cdot(1+x)\cdot\ldots\cdot(1+x).$$

The key point is the following: each of the addends that appear on working out this product arises from choosing from each parenthesis a 1 or an x in all the possible ways. So, to calculate the coefficient of the term x^i (that is, a_i) we proceed in the following way: we choose i monomials of the ns that there are in the product, from where we shall take the corresponding x, and from the rest of the monomials we shall take the 1. If we do this for every value of i, the binomial theorem turns out in the following way:

$$(1+x)^n=\binom{n}{0}x^0+\binom{n}{1}x^1+\binom{n}{2}x^2+\ldots+\binom{n}{n-1}x^{n-1}+\binom{n}{n}x^n.$$

From this relationship we can obtain other interesting ones: if we replace the variable x with a value equal to 1, we get the formula for the committee problem:

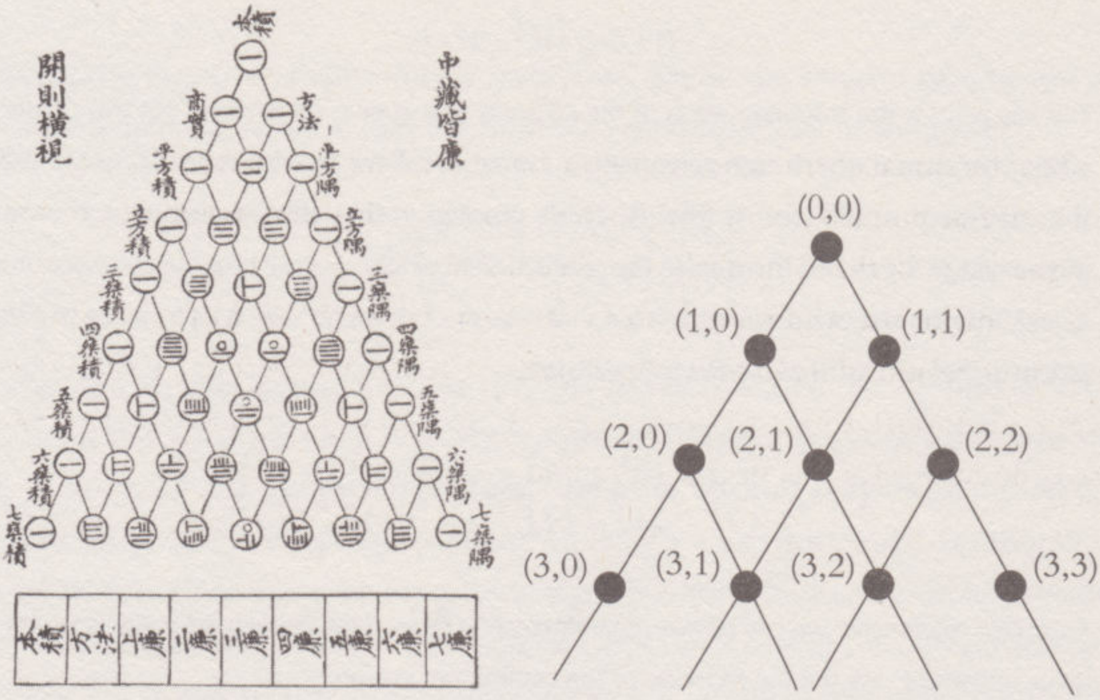
$$\binom{n}{0}+\binom{n}{1}+\ldots+\binom{n}{n}=(1+1)^n=2^n,$$

and if we write $x=-1$, we deduce the following equality:

$$\binom{n}{0}-\binom{n}{1}+\ldots+(-1)^n\binom{n}{n}=(1-1)^n=0.$$

There are many numerical relationships involving binomial numbers. One of them is what is known as *Pascal's triangle*, which also goes by the name of Tartaglia's triangle. Far from being merely a mathematical curiosity, Pascal's triangle can be

used to prove numerous combinatorial properties, in the same way that an abacus can be used as an instrument to carry out speedy arithmetical calculations. This is a construction that was already known in ancient civilisations. In eastern countries such as China, India or Iran this diagram had already been studied five centuries before Pascal stated its applications in 1653. In China it is known as the triangle of Yang Hui, in honour of the mathematician Yang Hui, who discovered it in 1303. The next diagram shows an oriental version of Pascal's triangle, and the simplification that we shall take to explain our arguments:



The Yang Hui triangle and the first vertices of a simplified model of Pascal's triangle.

The main point about using the diagram is this – the number of paths starting out from point (0,0), which we shall call *origin*, and which arrive at point (n,m), with the condition that at each point we go down one unit (to the right or to the left) is equal to the binomial:

$$\binom{n}{m}.$$

Note, in fact, that to reach point (n,m) we must make n steps in total, as each one allows us to go down one unit, or, equivalently, to increase the first coordinate

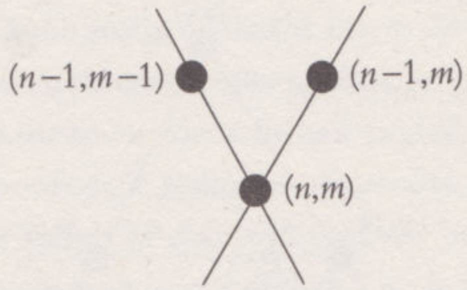
by one unit. Additionally, we have to carry out exactly m steps to the left as we go down the triangle, as the steps to the right do not contribute to increasing the second component. Therefore, the number of ways to go down from the origin to point (n,m) corresponds to the number of permutations with repetition, where the step to the left is repeated m times and the step to the right appears on $n-r$ occasions.

By using Pascal's triangle and its geometrical interpretation we can deduce combinatorial relationships directly, with no need to make any calculations. The first one is the value of the binomial:

$$\binom{n}{0}.$$

That binomial coefficient counts the number of ways to start off from the origin and arrive at point $(n,0)$. As we can only arrive at that destination in one unique way, the value of that binomial must be equal to 1.

Let's make the argument a little more complicated so as to obtain a more interesting relationship. To get to the horizontal level n we must first cross horizontal level $n-1$. In fact, to get to point (n,m) we must have beforehand passed through either point $(n-1,m-1)$ or through point $(n-1,m)$.



A detail of Pascal's triangle showing the paths leading to point (n,m).

By using the addition principle, we conclude that the number of ways of getting to point (n,m) is equal to the number of ways of getting to point $(n-1,m-1)$ plus the number of ways of getting to point $(n-1,m)$, or, expressed in the language of binomials:

$$\binom{n}{m}=\binom{n-1}{m-1}+\binom{n-1}{m}.$$

PLAYING WITH THE SYMMETRY OF PASCAL'S TRIANGLE

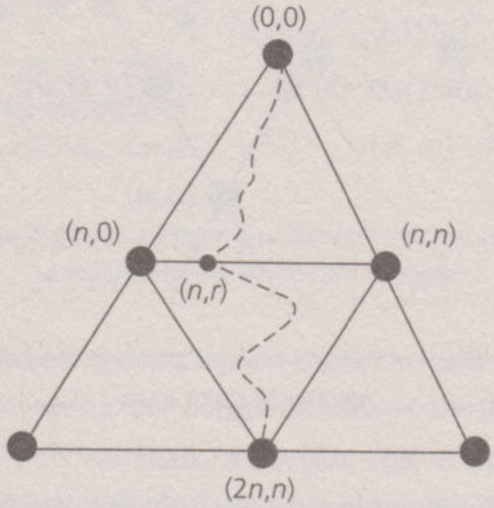
In the preceding cases we made use of the structure of the pathways in Pascal's triangle, but we can also utilise its symmetry. By applying symmetry in respect to the triangle's central axis, the paths that start from the origin and lead to point (n,r) become paths that start from the origin and lead to point $(n,n-r)$. Therefore, the following relationship of symmetry holds:

$$\binom{n}{r}=\binom{n}{n-r}.$$

This argument can also be applied to prove more complicated properties, such as the following:

$$\binom{n}{0}^2+\binom{n}{1}^2+\ldots+\binom{n}{n}^2=\binom{2n}{n}.$$

To prove it, we shall use the following property: we need to pass through some point of the form (n,r) to get from the origin to point $(2n,n)$. As the number of pathways from point (n,r) to point $(2n,n)$ is equal to the number of pathways from the origin to point (n,r) (note the figure's symmetry), the preceding formula holds by simply using the multiplication principle and the addition principle.



A pathway that starts from the origin and leads to point (2n,n) *, passing through level* n.

The reader is invited to try and prove this formula by using the binomial theorem which was outlined above in this chapter.

In this introductory section we have looked at the first concepts in the art of counting. Everything shown there will be very useful to us in later chapters, in which we shall study more complicated objects, such as graphs, maps, and trees, and will be fundamental in solving paradoxes and enigmas that stem from simple games of dice. In short, it will help in getting a more in-depth understanding of the notion of randomness.

Probability and combinatorics: two disciplines that go hand in hand

Many other disciplines feed off combinatorics. The calculation of probabilities in games of chance is an ideal channel to lead us into the basic concepts of enumeration. The multiplication and addition principles which we have introduced, together with their associated constructions such as binomial numbers will be of great help in understanding some of the curious paradoxes that games of chance throw up. That is because it is often better to know how to count accurately than to trust in the experienced gambler's instinct...

The origins of probability as a discipline came about as a result of a fruitful exchange of letters between Antoine Gombaud (1607-1684), who called himself 'Chevalier de Méré', and the mathematicians Pierre de Fermat and Blaise Pascal. Though Gombaud was not a member of the nobility, he adopted the title of a knight (de Méré, in reference to the town where he had studied) to sign his works. Besides being a fervent supporter of social justice as opposed to hereditary monarchy, one of his great interests was games of chance. Gombaud had made a note of certain paradoxes that arose in wagers on games of dice. This led him to write about these puzzles to the great French philosopher and mathematician Blaise Pascal, in the hope that a 'professional' could clear them up. This letter was the first of a series of correspondences between this shrewd gambler, the philosopher and, later, the lawyer and amateur mathematician Pierre de Fermat. A curious anecdote related to the story was this comment that Pascal made to Fermat in a letter dated 29 June 1654: "...de Méré is very talented, but he is not a geometrist; and that is, as you know, a very great defect..."

Despite the fact that we are not geometrists, either, we shall attempt to understand what probability is, and what a random phenomenon is. We shall talk about lotteries, games of chance in general and bets. It would therefore be a good

FERMAT AND PASCAL: THE LAWYER AND THE PHILOSOPHER WHO CHANGED MATHEMATICS

Pierre de Fermat's work as a lawyer in Toulouse did not prevent him from becoming one of the most influential mathematicians of the first half of the 17th century. Among his most important contributions are the first ideas for the algebrisation of geometry and the calculation of probabilities. However, it was in number theory where he was to earn himself the sobriquet of 'the prince of amateur mathematicians'. Without doubt, the best known anecdote about Fermat is the story of what he wrote in the margin of his edition of Diaphantus's *Arithmetica*: "(...) It is impossible to separate a cube into two cubes, or a fourth power into two fourth powers, or in general, any power higher than the second, into two like powers. I have discovered a truly marvellous proof of this, which this margin is too narrow to contain (...)". This problem would be solved 400 years later by using techniques that were way out of the reach of the lawyer, giving rise to what became known as 'Fermat's last theorem'.

Blaise Pascal's work spans a wide range of disciplines, from mathematics through physics and on to philosophy. Besides helping to set down the foundations of probability, Pascal also made important contributions to research in hydrostatics. As a result of a near-death incident and subsequently experiencing a religious vision in 1654, Pascal's religious beliefs became more entrenched, to such

an extent that some years later he would devote himself almost entirely to philosophy and religion. It was within this context of religion and mathematics that Pascal introduced what became known as 'Pascal's wager' in his work *Pensées*: Let's suppose that the existence of God is not known. If we were to bet on whether God exists or not, Pascal reasons that the best choice would be to bet that God exists. That is because, even though the probability that God exists is exceedingly low, such a low possibility would be amply rewarded by the great profit to be gained on achieving eternal life.

idea to use an abstract notation covering all these concepts so as to tackle the problem in terms that can be generalised. When a die is thrown, for example, there are numerous factors that will influence the result – the angle of the throw, its speed and moment of inertia, its height relative to the table, friction with the air... There are, therefore, difficulties inherent to the experiment that make it impossible to recreate all the factors that appear. All this complexity can be encapsulated in the notion of *randomness*.

Each time the experiment is carried out it has different initial factors and, therefore, it gives results that are completely unforeseeable on account of the complexity of the system. How can we measure the unforeseeableness of a result in an experiment? We can repeat it a large number of times and note down the results obtained and then see the proportion of times that the desired result appears. By means of this experimental method we can check the theoretical *frequency* with which a certain result happens. When we make the number of repetitions tend to infinity, the proportion that we obtain comes near to a theoretical abstract value which we call *probability* that the experiment will produce a determined result. Probability is, in short, an abstract concept that to a certain degree models going into the infinite in the process of counting frequencies.

How can we calculate those probabilities? The combinatorics of counting provides us with the following rule, commonly known as *Laplace's rule*:

"The probability of a determined event is obtained by dividing the number of favourable results that form the event by the number of possible outcomes of the experiment."

Laplace's rule allows us to cross over the border separating the world of randomness from the world of combinatorics. With this tool we can now attempt to answer the crucial question that Chevalier de Méré asked of Pascal. The gambler had studied the frequency with which events related to gambling with dice occurred. He had a specific enigma that he was unable to solve. He had noted that it was profitable to bet on getting at least one 6 when a dice is thrown four times, while betting on at least one double 6 on throwing the dice 24 times was not. A priori, this observation contradicts elementary mathematics, because 6 (the possible results on throwing a die) is proportional to 4 (the times the die is thrown) as 36 (the possible results on throwing two dice) is proportional to 24 (the times the two dice are thrown).

Pascal gave him the answer to the enigma by using the combinatorial techniques that we worked through above. Let's analyse the two situations mentioned by using Laplace's rule. In the first scenario, the number of possible cases comes to $6 \cdot 6 \cdot 6 \cdot 6 = 1{,}296$. The number of favourable cases is associated with those throws in which at least one 6 is obtained. Note that this value is equal to 1,296 minus the number of cases in which no 6 is obtained. This second value is easier to count and is equal to $5 \cdot 5 \cdot 5 \cdot 5 = 625$ (note that we have again applied the multiplication principle). Therefore, the probability of getting at least one 6 in four throws of a die is equal to:

$$\frac{1{,}296 - 625}{1{,}296} = 0.5177.$$

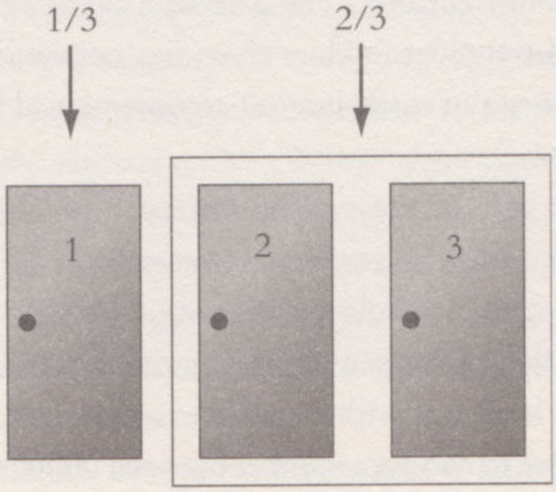
For the second situation the argument is similar. The number of possible outcomes is, once again by means of the multiplication principle, equal to 36^{24}, and the number of favourable outcomes is equal to $36^{24} - 35^{24}$. Therefore, the probability of success in this case is:

$$\frac{36^{24} - 35^{24}}{36^{24}} = 0.4914.$$

These calculations show that in the first scenario it is beneficial to bet, while in the second one it is not. The subtle difference in probabilities concerning even odds (50 per cent chance of winning or losing) was guessed by this shrewd gambler – in spite of the astronomical figures that are involved!

By using combinatorics techniques we can explain the following paradox which appears in probability, known as the 'Monty Hall paradox'. Fans of TV quiz shows may be familiar with this one. The dilemma first appeared in 1963 in the US television programme *Let's Make a Deal*, hosted by Monty Hall. Suppose that we have been invited to take part in the following game. We are shown three closed doors. We know that behind one of them there is a fabulous prize (such as a car), while behind the others there is nothing. Under these conditions and by following our instinct, we make a choice. Monty Hall, the compère, then opens one of the doors that has not been chosen and which does not hide the prize. He then gives us the option of changing the choice we made. What is the best choice to make in this situation, to stick with the initial choice of door or to swap it for the other remaining door?

Intuition tells us that the probability of winning or losing the prize does not change, as, once one of the doors without the prize has been opened there are two left closed and that, therefore, there are equal chances of the prize being behind one or the other. That argument is, however, incorrect, because the door Monty Hall opened never contains the prize – it is that point that holds the key to understanding that the equal chances argument cannot be correct. Let's analyse this problem by using combinatorics. Let's denote the doors by using labels – label 1, label 2 and label 3. Note that we can choose doors and win the prize in 3 different ways, while there are 6 different ways of not getting it right. Put another way, by Laplace's rule, 1/3 of the times we will win the prize, while more commonly (with a probability of 2/3) we will be unlucky and will not have guessed correctly.



In the Monty Hall paradox, the probability of guessing behind which door the prize is hidden is 1/3, while the probability of not doing so is 2/3.

Therefore, in 6 out of 9 occasions it is beneficial to change our choice after receiving the information that Monty Hall provides, while in 3 out of 9 we should not change our choice. The key factor is that Monty Hall opens a door with no prize, and we should make use of that information somehow. In probabilistic terms, it will always be a good idea to change doors if we want to win the prize!

As the reader may have guessed, in this game there is an interesting underlying question that goes beyond TV shows. Let's suppose that we want to know the probability of it raining today and, additionally, we know that we are in the rainy season. There is advance knowledge that gives us information on the event and which is conditioning it. That same fact can be extrapolated to the case of the Monty Hall show.

All this probabilistic theory on the conditioning of probabilities is the basis of what is known as *Bayesian probability*, named after the Presbyterian minister Thomas Bayes (1702-1761), who carried out the first studies into this subject. Those studies were the starting points for the fields of *Bayesian inference* or the *estimation theory*, which, starting from a set of data known beforehand, attempts to estimate what value is the most likely to appear next. These ideas are applied in fields as diverse as finance (in the estimation of the price of a share by studying its evolution during the previous weeks), the processing of signals (in digital signal recovery by the use of filters) or population dynamics (in the estimation of a determined population by knowing its historical evolution).

The Monty Hall show can rightly be called a game of chance. And, as we have seen, being able to count correctly can give us a slight advantage so as to make intelligent decisions. In popular folklore there is a number of beliefs about games of chance which, by mean of mathematical arguments, can be shown to be simply superstitions.

Let's consider the case of the weekly lottery. It is widely believed that it is impossible for the same result to come up two weeks in a row and that if by some chance it were to happen, it would be amazing. Nothing could be further from the truth; in fact, for that to happen or for any other predicted number to appear is equally likely! Let's look at it with a simple example: let's suppose that we play a lottery in which we have to choose a number between 00000 and 99,999. To be more specific, let's suppose we choose number 45,567. What are the chances of winning the prize? By applying Laplace's rule again, the probability equals

$$\frac{1}{100{,}000},$$

as the number of favourable outcomes corresponds to the choice we made, while the number of possible cases corresponds to the quantity of numbers between 00000 and 99,999, both inclusive. What will be the probability of winning the following week if we bet on the same number? The number of favourable outcomes and that of possible cases are still the same ones, so the probability will be the same! In other words, games of chance have no memory. In the field of probability, this notion is called *independence*, and what it really says is that two different goes at the same game have no influence on each other. This is the case of what happens in many

ROULETTE, OR HOW TO BREAK THE CASINO'S BANK

Roulette is one of the most popular games of chance in the world's casinos. It consists of a wheel that spins horizontally on its axis. The wheel's perimeter is divided into 37 spaces numbered, in no regular order, from 0 to 36, and painted red or black. The croupier spins a ball in the opposite direction to how the roulette is turning and after going round several times the ball falls into one of the numbers – that is the winning number. Roulette has numerous types of bets and combinations (only even numbers, only red, and so on). It is therefore impossible to predict the result of the experiment. The casino will, in fact, win in most cases.

There was, however, a Spanish family called the 'Pelayo clan' who became world famous for winning not inconsiderable amounts of money playing roulette in different casinos all over the world. Their strategy was based on the following idea: as the roulette table is a physical object, it must have some imperfection, such as an axis giving an imperfect spin, or the wheel being slightly unbalanced. So, if these defects can be detected empirically, then it will be possible to work out situations whose outcomes will have slightly higher probabilities than according to the theory. By making use of this idea they became the scourge of gaming houses round the world. Unfortunately for gamblers nowadays, roulette tables are now regularly changed round or replaced, so such strategies are no longer effective.

commercial games of chance. Not only do they behave with independence, but the preceding arguments show that there are no combinations more likely to provide prizes than others. In the lottery, a number such as 00001 is just as likely to win the prize as a more attractive number such as 48,756. Probability shows no preference for numerical aesthetics!

The natural impulse to take part in a determined game of chance is prompted by the financial prize to be awarded to the winner. Before gambling with our hard-earned money, we should weigh up whether the outlay is in accordance with the probabilities of winning, and whether the uncertainty regarding winning the prize is worth the possible economic loss. In plain speech, we gamble because there is a possibility that we will win. There are cases in which it is more advisable to gamble than in others, and to differentiate between them we need more information than just the probabilities: we shall need what is known as *mathematical expectation*.

Let's look at the following game which, though not terribly exciting or entertaining, will enable us to introduce the notion of mathematical expectation

in a simple way. Suppose we have two perfectly balanced dice, in other words, the probability that a determined value comes up in one throw is 1/6. We will be given £4 if, on rolling the dice, we get the same value on each, while we lose 1 pound if the values are different. At first sight it does not seem like the person offering to play the game with us is trying to trick us into losing our cash, as, when we win we receive a lot of money, while when we lose we only have to give that person a small amount. To work it out, let's note that of the 36 possible combinations that can be obtained on throwing two dice (6 combinations for the first die and 6 combinations for the second give a total of $6 \cdot 6 = 36$ possible outcomes), only 6 of them produce two equal values, while the rest give values that are unequal: on 6 out of 36 occasions we will win £4 , while in the rest we will lose 1 pound. So the expected amount we will win is obtained *by weighting* the outcome by its probability. In our case it will be equal to:

$$4 \cdot \frac{6}{36} + (-1) \cdot \frac{30}{36} = -\frac{1}{6} = -0.16666,$$

and, therefore, it is in our interest not to gamble, as the value obtained is negative. However, if, instead of winning £4, we could win £30, the expected value would become:

$$30 \cdot \frac{6}{36} + (-1) \cdot \frac{30}{36} = \frac{150}{36} = 4.1666,$$

and in this scenario it would be in our interest to play, as the high-value prize for winning compensates for the lower likelihood that two equal values appear in throws of dice.

A consequence of all these ideas is shown in the next example using the classic rules of the EuroMillions lottery. In this lottery (in which the bet costs £2) you need to get 5 numbers right out of 50 possible numbers, as well as 2 numbers from 9 additional possibilities (the so-called Lucky Stars). In any case, we are dealing here with choices without order (the sets matter, but not the order in which they were chosen), and therefore, by applying the results relative to the enumeration of structures without order, we find that the number of possibilities is equal to:

$$\binom{50}{5}\cdot\binom{9}{2}=76{,}275{,}360.$$

The probability of winning a first prize will, therefore, be 1 in 76,275,360; very small indeed. There are no magic formulae for winning the lottery, and the truth is that the chances of winning this type of game are usually minimal (as the figures show). Nevertheless, there are times when it is worth gambling a few pounds for the chance of improving one's standard of living considerably.

In November 2006, the EuroMillions jackpot came to £124 million. The chances of winning the first prize were really small, but the accumulated jackpot was an encouragement to bet. We shall work out the mathematical expectation for this gamble with those values. Bearing in mind the number of possible cases (76,275,360) and the number of favourable cases (1), it turns out that the expected number of pounds comes to:

$$180{,}000{,}000\cdot\frac{1}{76{,}275{,}360}-2\cdot\frac{76{,}275{,}359}{76{,}275{,}360}=0.3598708941,$$

THE MATHEMATICAL DEFINITION OF EXPECTATION

Let's get back again to mathematics and language. Suppose that in a certain random experiment we can get N different numerical results $x_1, x_2, \dots, x_N$ where the probability that we get the result x_1 will be p_1; the probability we get the result x_2 will be p_2, and so on up to N. The expected value of our experiment (or mathematical expectation) is obtained by averaging each of the possible values by the probability of obtaining that value. In other words, the mathematical expectation is the following sum:

$$p_1x_1+p_2x_2+p_3x_3+\ldots+p_Nx_N=\sum_{i=1}^{N}p_ix_i.$$

Note that if all the cases have the same probability of occurring, then $p_1=p_2=\ldots=p_N=1/N$, and the expected value will be the arithmetic mean of the values obtained. To sum up, we should understand mathematical expectation as an arithmetical expectation (which provides information that is partially representative of a determined experiment) in which we give weight to different values according to their probability of appearing.

which is a positive value. In this scenario it was worth spending £2 to try your luck on account of the astronomical amount of prize money. Despite the fact that mathematicians have very clear ideas regarding bets, and we often hear the saying "the lottery is a voluntary tax imposed on those who don't know anything about mathematics," on this occasion there was a certain advantage for those who knew how to make ingenious calculations...

Mathematical expectation is a tool that gives a description which is only partial, but it often provides sufficiently useful information so as to be able to understand the phenomenology of the event being study. In fact, as we shall show in the following chapters, averages and probabilistic arguments will provide us with much more information than it would seem. They will open the doors to the intimate relationship that exists between combinatorics and probability. And this will allow us to prove surprising results with subtle techniques and ingenious ideas.

Chapter 2

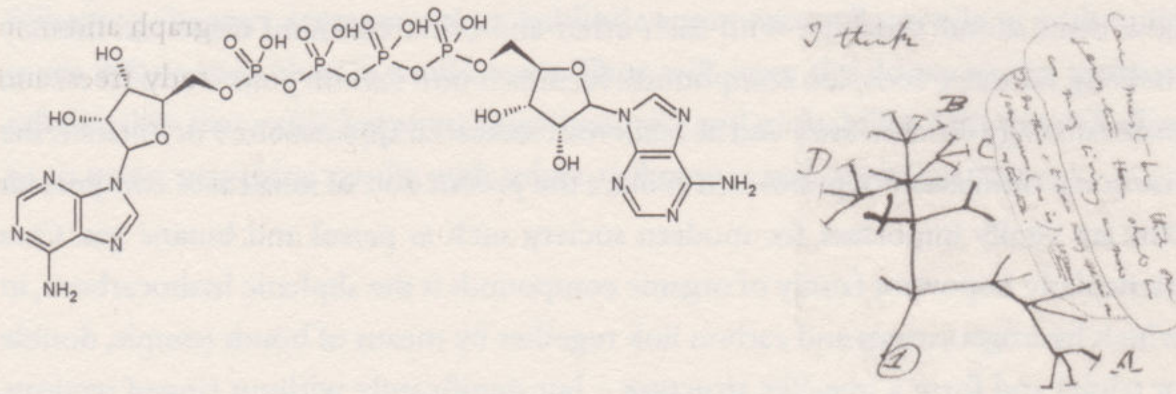
Graphs and Maps

The fundamental element for life is carbon. Every living creature contains carbon, from the most elementary bacteria to the most gigantic mammals. Given its relevance, a whole discipline of chemistry – organic chemistry – is involved in studying how those atoms combine with each other and other chemical elements, thereby forming far more complex compounds. Research into carbon plays a crucial role in pharmacology, biochemistry and in numerous industrial applications. For instance, the fractional distillation of petroleum enables the production of numerous compounds that are vitally important for modern society, such as petrol and butane gas. One particularly important family of organic compounds is the aliphatic hydrocarbons, in which hydrogen atoms and carbon link together by means of bonds (simple, double or triple) and form a tree-like structure – but significantly without ringed sections.

Let's completely change the subject. Everyone has heard of the theory of evolution developed by Charles Darwin from observations he made while voyaging all over the globe on his expeditions on the *Beagle*. Years later, that analysis would engender *On the Origin of Species*, a book which caused quite a controversy because its thesis was outrageous for that era. That work contains something that is very interesting for our studies – the form known as the *tree of life* showing the different species. The construction of this complex tree is made up in the following way: groups of different species are formed, which are made up of the pairs of species that are most similar (under certain criteria of similarity), and the process is repeated successively until all possible species are covered. By means of this process a *phylogenetic tree* or *taxonomy cladogram* of the species can be built. This phylogenetic tree shows the degree of relationship (or of metrics) between species that are similar to a lesser or greater degree. These phylogenetic techniques can be applied, for instance, to studying the degree of relatedness of numerous species of animals from a determined geographical region, or to showing the different evolutionary steps that led to *Homo sapiens* being as it now is.

The reader may find it a little puzzling that in a book on mathematics we should be talking about organic molecules and phylogenesis. And even more so if these

topics are from such disparate fields as chemistry and the theory of evolution. To set the reader's mind at rest, what we shall show is that these two fields (as well as many others) and that of combinatorics have much more in common than might initially be imagined. Mathematics is, much of the time, abstract concepts of real objects, and that is what occurs in the two scenarios mentioned previously, since the structures that define the molecules of saturated hydrocarbons and of taxonomy cladograms are essentially the same.



The structure of an organic molecule and a drawing of a phylogenetic tree by Charles Darwin. The two have a very similar look to them.

Observation of the figures above allows the conclusion to be drawn that both have a structure that is linear and *ramified*. This is the basic notion behind understanding what a combinatorics tree is. For a combinatorialist, a tree is neither a molecule nor a plant but rather an abstract structure with numerous applications in real life (chemistry, taxonomics, etc.) and in theoretical IT (decision-making algorithmics and databases) that has mathematical properties which are useful and fascinating in themselves.

Faced with a particular phenomenon, and armed with an intuition based on data from experiments and reasoned arguments, we can construct abstract models that clarify at least part of the question. The most curious thing about it all is that problems and issues in areas that are completely unconnected with each other so often provide identical mathematical models. This is what occurs in the two examples we are considering. This is a rather common feature of combinatorics, especially in graph

theory. Relationships between individuals, social networks, the spread of epidemics, chemical molecules, supply networks and the like can be modelled by using the discrete structures that we shall study next. We shall start with the definition of *graph*, a fundamental concept in the field of discrete mathematics and combinatorics. As we shall see, that concept is purely set-theoretic. Later, we shall see what properties appear when we draw a graph on a piece of paper and obtain a more rigid structure called a *map*. That was introduced by one of the greatest specialists in combinatorics in the 20th century, William Tutte. The aim of defining these objects formally was to carry out a more rigorous study of one of the most important problems of the 20th century – the four-colour problem. Later, once the notions of graph and map have been covered and understood, we shall go on to define and study trees, and to see how those objects are really key factors in gaining a better understanding of more complicated combinatorial objects. On the way we shall discover ancient Tibetan games, problems on triangles and nature's hidden numbers. Everything is mathematical.

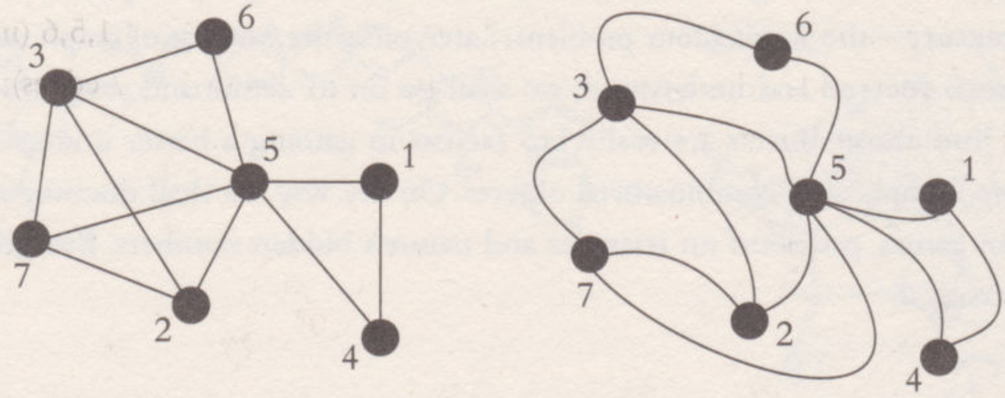
Graphs and maps

The fundamental topics in this chapter will be graphs and maps. Before going on to define them with precision, we shall first provide a basic idea of their concepts. Let's imagine the place where we live. Irrespective of whether we live in a great city or in the countryside, it is very likely that in the area where you live there will be a number of small towns. And as a result of progress in industry and communications, roads will have been built to link the towns. It might be that to go from one town to another we can do so directly, or that we have to go through some other town, or we might even have to use a flyover to cross over another road. There is, therefore, a notion of connectivity between the towns (if we can travel directly from one to another), and a notion of planarity (that is, unless roads crossing is inevitable and bridges have to be included in the road system).

Graphs condense the idea of the connectivity mentioned above: a labelled graph is a set-theoretic structure formed by *vertices* (or points), which carry a label allowing them to be differentiated. There are also relations of incidence between these vertices. The usual way to represent graphs is by drawing the set of vertices on the plane (with their corresponding label) and by drawing a line or *edge* that joins two vertices if and only if the vertices are incidental. That is why a graph G is usually written

as $G = (V,A)$, where V is the set of vertices of the graph and A is the set of edges. Note that the way the edges are drawn does not matter; what interests us is which vertices are connected.

It is also common to use the term *unlabelled* graphs, which are obtained from labelled graphs by overlooking or deleting the labels and keeping the skeleton of vertices and edges. Depending on the different contexts, the reader will understand whether we are referring to labelled graphs or to unlabelled ones.



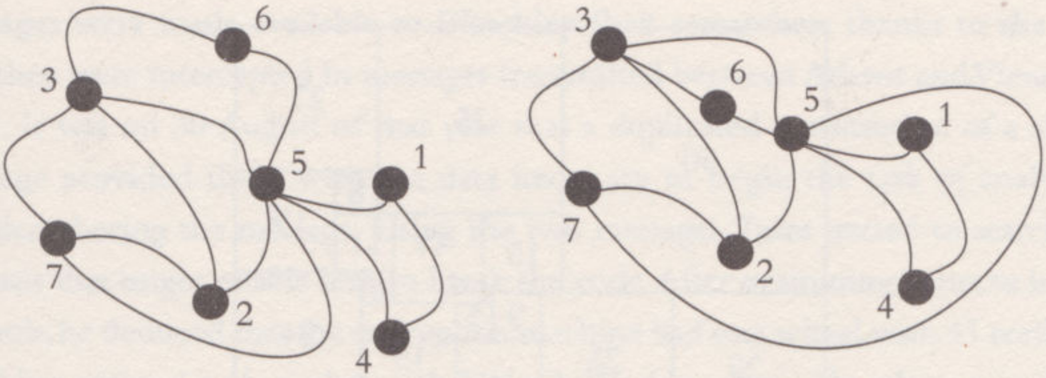
The same graph shown in two different ways.
Note that the connections between the vertices are the same in both cases.

Note that the figures above show the same graph in ways that are qualitatively different. In the first one straight lines were used, while in the second we used curved lines. It is also important to note that in the first figure the edge defined by vertices 5 and 7 crosses the edge defined by vertices 2 and 3, while the second one shows the same graph without intersections of edges. Thus, if the first graph were used to model a road system then we would have to include a bridge, though in actual fact it should not be necessary as it is possible to implement a different system which avoids all crossings.

Such representations of graphs without crossings on the plane are interesting in themselves, and have their own name – a *map* is a representation of a graph (whether labelled or not, depending on the context) on the plane, in such a way that there are no intersections between the edges. Drawing a graph is what in mathematics is known as *immersion*, a concept that models the notion of drawing an object inside another object. The first figure of the two given above is not a map, but the second one is. In fact, one and the same graph can represent diverse maps, all corresponding

to different immersions of the graph on the plane. That is due to the fact that the graphs are composed of vertices and incidences (or intersections) between them (the edges), while maps have the additional structure of regions in which the plane is subdivided. Those regions are what we call *faces* of the map. Maps, therefore, are structures which are more rigid than graphs, a fact that means that the study of them is often more multifaceted.

We shall show an example to clarify these ideas. In the next figure one and the same graph is shown embedded in the plane in two different ways. In fact, the two immersions give rise to two different maps. In the first one, the external face (also called the unbounded face) is defined by the vertices 6,3,7,5,4,1,5,6 (in anti-clockwise order) and is, therefore, a face of length 7 (it is defined by 7 edges); in the second immersion, however, there is no face with that length.



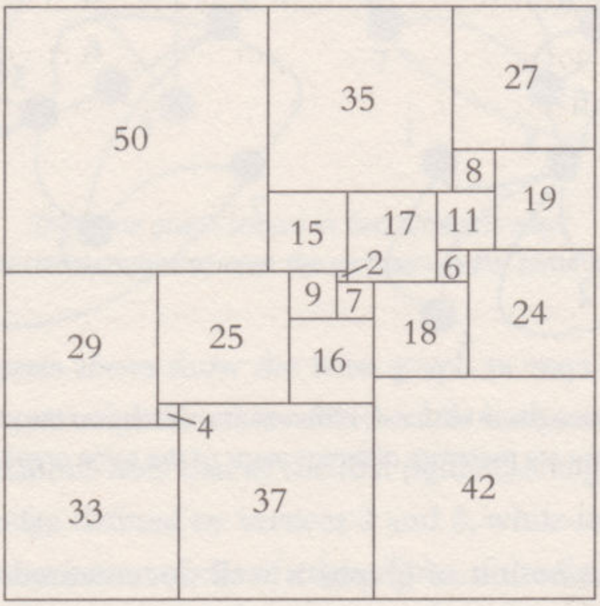
The same graph with two different immersions on the plane; they are therefore different maps of the same graph.

The origin of the notion of graphs is well documented. It first appeared in relation to the famous problem of the Bridges of Königsberg, which was solved in 1736 by Leohnard Euler. But the origin of the notion of a drawn graph, or map, is more recent and at the same time less well known. Though the first formal definition of map was made by the mathematician Jack Edmonds, the first person to carry out an enumerative study of map theory was a chemist attracted by the beauty of mathematics who played an important role during World War II as a cryptanalyst for the Allied forces. His name was Bill Tutte.

William Tutte was born in Newmarket, England, in 1917, the son of a gardener and a housewife. During his primary schooling he already showed a certain aptitude for science, particularly in the field of chemistry. Following his secondary studies he

went on to study chemistry at Cambridge University. Nevertheless, his interest in mathematics only increased and he often attended meeting of mathematicians at the illustrious Trinity College of Cambridge. It was at that time and during those heated debates where Tutte, along with three other young mathematicians, Roland Brooks, Arthur Stone and Cedric Smith, solved the problem known as *Squaring the Square.*

This problem goes as follows: what integer lengths can the side of a square have in such a way that it can be sub-divided into different squares? Though the Trinity four were still only students, they managed to solve the problem by using arguments related to electrical networks. Later, these ideas evolved into what is known as *flow theory* in graphs. Their teamwork was to continue, and the four researchers went on to work together on solving problems, signing their work under the pseudonym *Blanche Descartes*.



A valid squaring of the square: all the smaller squares are different.

At this point in the story, World War II had already begun, and Bill Tutte became immersed in important chemical research. His university tutor realised that his ability in mathematics meant that he would be an important talent for the British military, who could make good use of his potential for deciphering enemy codes. He was sent to Bletchley Park, also known as Station X, together with some of the greatest mathematicians of his generation, and given the task of intercepting and breaking

the German force's codes. It was to be at that centre where he managed to break the code known as *Enigma*, probably the most famous code from World War II and the subject of many articles and books. So, it came about that in the January of 1941 Bill Tutte started work as a cryptanalyst in the service of the Allies. His work as a decoder was impressive. On being appointed Officer of the Order of Canada in 2001, it was said:

> "...As a young mathematician and code breaker, he deciphered a series of German military encryption codes known as FISH. This was considered to be the greatest intellectual feat of the Second World War...".

There are even today many details of those codes that have still not been made public on account of the military secrecy surrounding them. The first FISH messages were made available to Bletchley Park researchers thanks to the fact that they were intercepted in messages transmitted between Athens and Vienna in 1941. It was on 30 August of that year that a duplicated transmission of a single message provided them with the data necessary to begin the task of analysing and deciphering the message. Using the two messages, Tutte started to search for patterns that might enable him to break the code. After examining patterns in the symbols, he deduced that the encryption machine had one wheel with 41 teeth (or possible positions) and another with 31 teeth. Working alongside other researchers, he finally reached the conclusion that the machine had a total of 12 wheels, and they gradually uncovered the complete cipher. Bearing in mind that they had only a small number of coded messages to go on, what a titanic task it was!

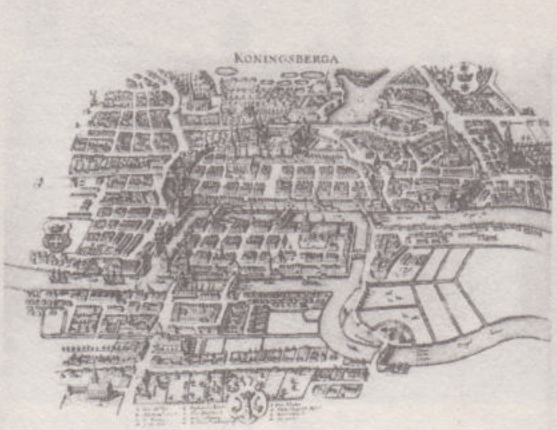
Bill Tutte and the entrance to Bletchley Park.

Besides studying the internal structure of the FISH codifiers, Tutte wrote algorithms for deciphering their codes. These routines would be successfully put to use a few years later with Colossus, a computer designed specifically for breaking the enemy's codes. Note that this was in the 1940s, so Colossus was therefore the great-grandfather of computers as we understand them today. All of this boils down to the fact that the intellectual activity at Bletchley Park was truly amazing, and research carried out there would subsequently be of vital importance in encryption, in algorithmics and in the design of computers. Despite the fact that many researchers received recognition and honours for their work there (as in the case of Alan Turing) Tutte never received public recognition for his work as a cryptanalyst.

Recognition was to come years later in Canada where he specialised in graph theory and combinatorics. It was in this context that he turned his attention to the four-colour problem, attempting to solve it with a purely enumerative strategy. Remember that this problem consists of the following: how many colours on a map are enough to colour its vertices in such a way that adjacent vertices never have the same

THE ORIGIN OF GRAPHS: WALKING AROUND KÖNIGSBERG

It is usually impossible to determine the exact time at which a new scientific discipline is created. But that is not the case with graph theory. This field of mathematics has a very specific date of birth. In 1736 in the Prussian city of Königsberg (now known as Kaliningrad in Russia) there was a puzzle that appeared to be unsolvable: no one had been able to find a way to walk around the city, starting off from their home and returning there having crossed each of the bridges over the Pregolya River just once. It was Leonhard Euler who put forward the idea of creating a simplified model of the city out of vertices and edges. The edges represented the bridges connecting each of the parts of the city. By means of simple arguments on the graph he had defined he was able to demonstrate that the route people sought did not exist and, at the same time, Euler spawned a completely new branch in the world of mathematics.

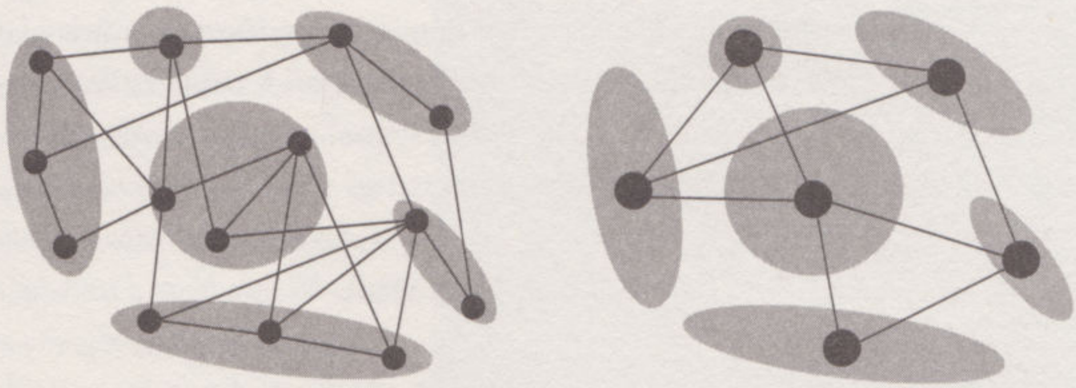


The city of Königsberg, with its seven bridges, in 1652.

colour? Proving that it can always be done with five colours is an elementary exercise, but deciding if the minimum is four or five is an extremely complex problem, which was not solved until Kenneth Appel and Wolfgang Haken managed it using computers in 1976. Although Tutte did not find the answer to the problem he attempted to count how many maps can be coloured with four or with five colours so as to decide if the counts were equal or different. This research made Tutte a pioneer in the use of enumerative techniques in mapping.

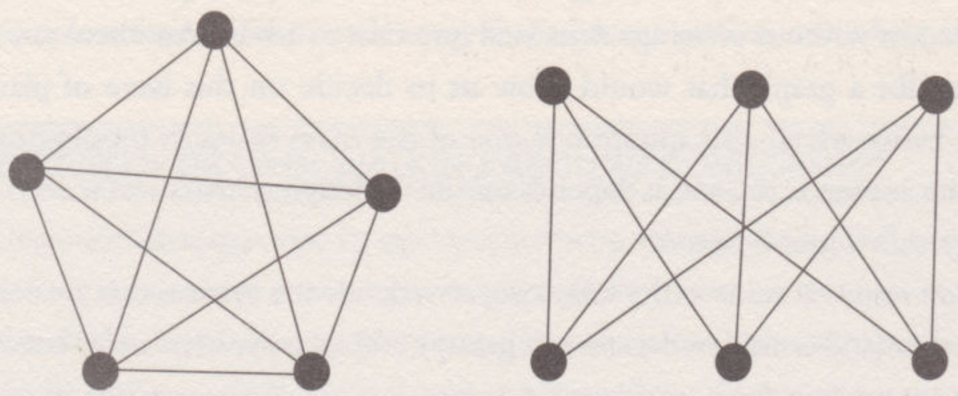
Let's go on to analyse the close relationship between graphs and maps in more detail. A fundamental question regarding that relationship is knowing when a graph can be represented as a map, in other words, when a graph can be drawn on the plane without the need for crossings. If that happens, we say that the graph is *planar.* We have seen already how the first example given was not well thought through as it had crossings, but the second showed to was possible to redraw the graph on the plane without any crossings. Therefore, that graph is planar, as it admits a representation without crossings. A natural question to ask is: "Are there any *intrinsic* conditions for a graph that would allow us to decide on this issue of planarity?" Far from being trivial, that question is one of the main issues in topological graph theory. The answer is yes, and it depends on the underlying structures within a graph which are called graph *minors.*

A graph minor is made in the following way: divide the vertices that are connected to each other (and considered to form a group), and take the incidences between the different groups (see the next figure). A minor is a graph, the vertices of which are the groups (the dark parts that contain the vertices), and the edges are the incidences between the groups as a whole. The figure below shows the construction of a graph minor in such as way.



A graph and one of its minors.

In the world of graph theory there are certain graphs that are rather special, and which, on account of their importance, are given their own names. That is the situation with the problem we are dealing with here. Our celebrity graphs in this case are the two shown in the following figure: the graph with five vertices and all the possible edges (the graph denoted by K_5 which is called a *complete five-vertex graph*) and the graph with six vertices and nine edges (denoted by $K_{3,3}$). The reader can try to draw these graphs on the plane without crossings and will find that it is impossible. They can also try to show that if, on either of them one of the edges is eliminated. In this case, it is possible. These two graphs are, in fact, the smallest (as far as the number of vertices are concerned) with a non-planar property. In this sense they are maximal, as when any edge is eliminated they take on the property of being drawable without crossings.



The complete five-vertex graph, K_5*, and the bipartite graph* $K_{3,3}$.

By using elementary arguments, it can be proved that these two graphs are not planar. But what is really surprising is that the reciprocal result is similar. If a graph does not have either of the above-mentioned graphs as a minor, then it is planar. This now classical result is due to the Polish mathematician Kazimierz Kuratowski (1896-1980), the greatest exponent of the Polish school of mathematicians in the first half of the 20th century *Kuratowski's theorem* states:

"A graph is planar if, and only if, it does not contain $K_{3,3}$ or K_5 as a minor."

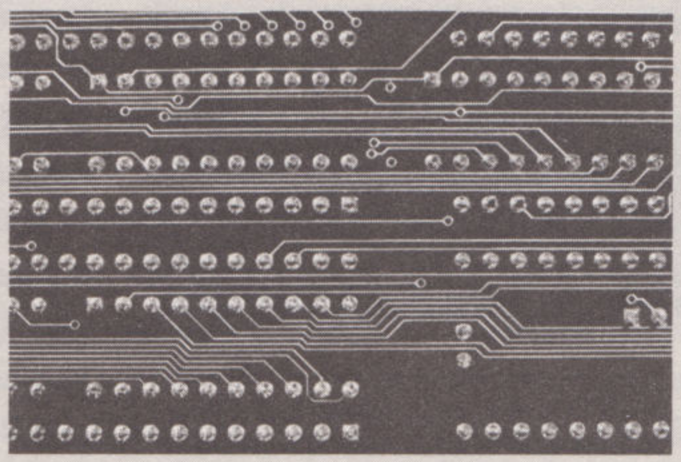
It is very interesting to note the following. A topological condition such as planarity (which depends on the surface on which we are drawing the graph) can

be defined only in set-theoretic terms (as the condition of being minor is merely a relationship of incidences) and, therefore, that property is *intrinsic* to the graph, and not to how it is drawn on the plane. The second point to be emphasised is that such a complicated condition as it can be to design a highly complex graph without crossings becomes a search for just two well-defined sub-structures existing within it. Kuratowski's theorem is, in fact, just the tip of the iceberg of an extremely complex and rich mathematical construction which is focusing the attention of the majority of specialists in combinatorics and the study of algorithms within what is known as *graph minor theory.*

The family of planar graphs is said to be closed under minors, which means that any minor of a planar graph is also planar. This observation stems directly from the definition of minor which we gave previously. In this case, the Kuratowski theorem assures us that the graphs in that family are characterised by the two graphs that

AN INDUSTRIAL APPLICATION OF PLANARITY

Discovering whether a graph will allow a planar representation is a problem of great importance in industry. In the field of electronics it is necessary to know how to implement a given electronic model in the simplest way possible. That is because higher levels of complexity bring with them more costly production methods. The crux of the matter is that once the type of circuit model to be used has been decided, the next task is to find out if it can be represented on a planar printed circuit board. If the circuits cannot be made without crossovers the manufacturing process will be more expensive, as the silicon chips will need more than one layer to hold the various copper connections. Bearing in mind that such circuits are produced in millions of units, the study of planarity becomes a key factor for optimising costs. There are, in fact, various software programs designed exclusively for discovering whether an electronic circuit will allow a planar representation or not, and if it does, then to trace out its optimal representation for a printed circuit.



obstruct the condition of being planar, and no more. Is it true that, in general, any family of minor-closed graphs is characterised by a *finite* set of forbidden graphs? This question, called the Wagner conjecture, is a difficult problem, but the answer is yes. Any minor-closed family is characterised by a *finite* set of forbidden minors. That work took up a large part of the graph theory of the second half of the 20th century. In their titanic series of more than 20 scientific works (all entitled *Graph Minors* and numbered consecutively with Roman numerals) Neil Robertson and Paul Seymour laid the foundations for this mathematical theory. Their implications are tremendously important for present-day graph theory and, in fact, a large part of the research into this discipline is nowadays inspired by that monumental work. As a consequence of their work, there exists the following surprising result, named the *Robertson-Seymour theorem*:

"In any *infinite* set of graphs, there are two graphs in the set such that one is a minor of the other."

There are very profound implications in all this theory regarding the study of the structure of graphs; in actual fact, it is one of the most active topics of research in the field of combinatorics and algorithms in graph theory.

Some counts that are more complicated

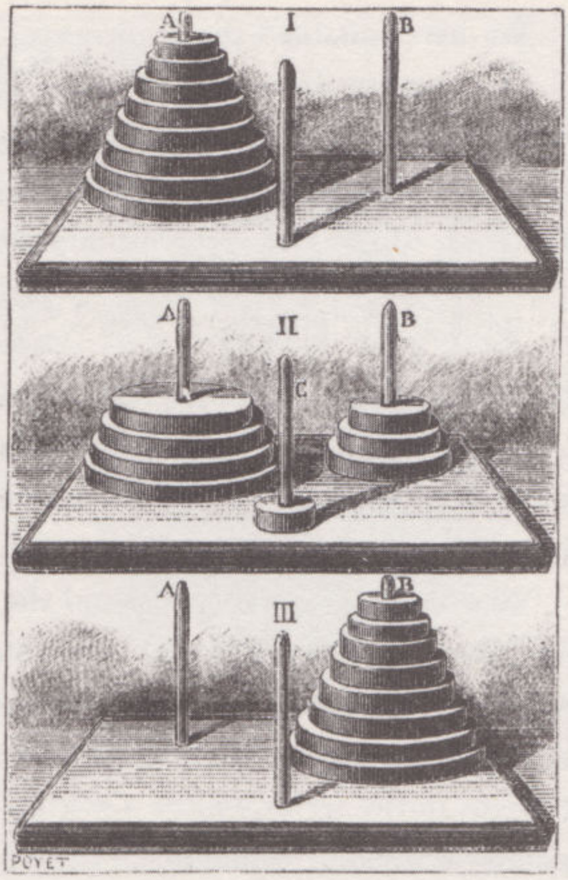
To be able to get a better and greater understanding of the enumeration of certain families of graphs and maps it is best here to make a pause so as to introduce some new enumerative techniques, which will allow us to deal with combinatorial situations that are more complex than those studied so far. To do so, we shall make use of an example that is very simple but at the same time extremely representative.

Remember that we introduced the notion of permutation, and we looked at how to calculate the number of permutations of a set of n elements. That value is equal to $n!$, which is defined as the product of the integers lower than n, going from n to 1. Let's imagine that we want to calculate the value of $120!$ and, therefore, we have to calculate the product of all the natural numbers between 1 and 120. In the same way, if we wanted to calculate the value of $119!$ we would have to realise a similar product, but this time without the term 120. Note that the following relationship holds: $120! = 120 \cdot 119!$. What are the implications of this? If we have

already calculated the value of 119*!* and we have it written down, then the value of 120*!* can easily be calculated by making use of it. In other words, we can take advantage of the *recursive* quality of the formula of the factorial to carry out the calculation by using other smaller, partial calculations without needing to work out the formula to the end every time.

The same situation appears on numerous occasions in combinatorics at different levels of complexity. Let's look at another example so as to clarify these ideas. We are going to count the number of words that can be written using the letters a and b, and which are of length n. Let's denote this number using P_n. Such a count will be defined by the variations with repetition, and the number of words with length n is equal to 2^n. Let's write $P_n = 2^n$. Now observe that any word of length n is built starting from a word of length $n-1$ by linking either the letter a or b to the end. This gives us the recursive relation $P_n = 2 \cdot P_{n-1}$, which shows the problem's underlying combinatorics. These arguments take us to another type of equations, those called *recurrence equations* , or recursive equations. In the case given, it would not have been necessary to use the recurrence equation in order to solve the problem, but in other ones it would be absolutely essential.

Let's look at a more complex case. The Tower of Hanoi Puzzle (also known as the Tower of Brahma game or The End of the World Puzzle) was invented in 1883 by the French mathematician Édouard Lucas (1842-1891). The game is based on an ancient oriental legend regarding the end of the world. To test the mental strength of new priests entering a sacred temple, the elders used to set them the following task: 64 rings of decreasing size are stacked up in order, as shown in this illustration. The objective of the game is to move all the rings to another place, with the condition that in each move only one ring can be moved



The initial position of the Tower of Hanoi puzzle, with eight rings, an intermediate position and the final position.

at a time. Furthermore, the player is not allowed to place a ring on another one of a smaller size. Players are allowed to use a middle position so they can carry out intermediate movements.

The objective is to find the minimum number of movements needed to complete the task. By using a simple argument we can answer the question: let T_n be the number of movements needed when we have n rings. The key is the following observation: to move n rings from the initial position to the final position, what we must do is first to move the $n-1$ upper rings to the intermediate position, to next move the largest ring to the final position and, to finish, to move the $n-1$ rings that are in the intermediate position. So, the number of moves needed equals $T_{n-1} + 1 + T_{n-1}$. We need T_{n-1} movements to move the $n-1$ upper rings to the intermediate position (in these steps we leave the largest ring in its place), one movement to put the largest ring in its final place and, lastly, T_{n-1} to move the $n-1$ upper rings to the final position. Note that the key is in forgetting about the largest ring, which does not cause any problems as we can always put smaller rings on top of it.

We have obtained the recurrence $T_n = 2T_{n-1} + 1$, with the initial condition $T_1 = 1$ (the game for one ring is trivial and is based on one single movement). Let's look for a formula for the value of T_n. By applying it we get the sequence of values 1, 3, 7, 15,... Examination of the first values obtained leads us to conjecture that $T_n = 2^n - 1$. And so it is, as this formula is true for $n = 1$, and additionally it holds that:

$$2^n - 1 = 2\,(2^{n-1} - 1) + 1,$$

giving rise, therefore, to the proposed recursive formula: $T_n = 2\,T_{n-1} + 1$.

By applying this formula for a value n of 64 we get the result that the young monks have to make a total of $2^{64} - 1 = 18{,}446{,}744{,}073{,}709{,}551{,}615$ movements... not a task for the faint hearted!

To conclude this stroll around the world of recurrence combinatorics we shall mention a classical problem which was studied back in the times of the Renaissance. Let's start off from the elements 1, 1 and build a numerical sequence by adding the two last elements that we get from the process. In this case, the third element in the series will be 1 + 1, making 2; the fourth element will be 1 + 2, making 3; the fifth would be 2 + 3, making 5; and so on. The first terms obtained using this rule are the following:

1, 1, 2, 3, 5, 8, 13, 21, 34...

This sequence is famous in the world of maths and goes by the name of the *Fibonacci Sequence*, in honour of the first person to study it, Leonardo of Pisa, the son of Bonaccio (the derivation of his moniker Fibonacci), an Italian scholar who lived in the Middle Ages. The specification by means of a recurrence for this sequence is as follows: we denote with F_n the n-th term of the sequence. Then the following recursive relationship holds:

$$F_n = F_{n-1} + F_{n-2}$$

It is curious to look at the context that inspired Fibonacci to come up with this numerical sequence. The key issue stems from the following problem concerning population dynamics: a pair of rabbits, starting from the second month of their lives, have a pair of rabbits per month, which, from their second month, also have a pair of rabbits, and so on. How many pairs will there be after a certain time? The Fibonacci

ÉDOUARD LUCAS AND THE FIBONACCI NUMBERS

Édouard Lucas (1842–1891), was a French mathematician who spent a large part of his scientific career in Paris, first in the city's observatory and later in two institutes – the Lyceé Saint Louis and the Lycée Charlemagne. His best known works in mathematics involve the study of a family of recurrences with close links to the Fibonacci sequence. Lucas noted that many properties of this sequence were preserved if the initial conditions were changed (remember that in the Fibonacci sequence we took 1, 1 as the first values of the series). There is, in fact, a numerical sequence named after him, the Lucas numbers, which are defined in the same way as the Fibonacci numbers but this time starting with the initial conditions of 2, 1.

The importance of the Fibonacci numbers and their generalisation goes far beyond simple anecdotes and there is even a scientific magazine, *The Fibonacci Quarterly*, which specialises in results related to this most distinguished of sequences.

sequence is a model for this puzzle. More important than the sequence in itself is the fact that it grows indefinitely (the sequence is defined in terms of a sum of positive whole numbers). The quotient of the consecutive numbers is never the same value, but rather tends to a limit value:

$$\frac{13}{8} = 1.625, \ \frac{21}{13} = 1.615384615, \ \frac{34}{21} = 1.619047619.$$

The successive quotients of the terms of this sequence approximate a special number, the so-called *Golden Ratio*, *Golden Section* or *Divine Proportion*, the value of which is:

$$\frac{1+\sqrt{5}}{2} = 1.618033988...$$

It was this factor regulating the growth of the family of lagomorphs that initiated Leonardo of Pisa to the study of the properties of this numerical series. The special proportion does, in fact, appear in many other contexts – in the way in which plants develop stems and in the proportion in which spirals grow on snails, as well as the increase in animal populations.

Nature behaves in accordance with certain rules, and one of them is the growth of biological structures under well-defined patterns. It could be said that nature chose this number on account of its harmony and perfection, though there are other more prosaic and complicated reasons of geometric character that we shall not deal with here.

It is precisely because of that harmony and balance that numerous artists, architects and musicians use what is known as the divine proportion in their creations. We could list a great number of architectural creations (such as the Parthenon in Athens), of articles of daily use (ID and credit cards) and other objects created by human beings in which that aesthetic harmony has been put to use, but here we shall just show the painting named *The Third of May*, by the great master Francisco de Goya. In the painting, the main character – the man about to be executed – can be seen at a point in the picture that defines four rectangular regions whose sides are of proportions linked to the divine proportion.

Aesthetic rules and patterns are very often governed by the divine proportion. One good example is Goya's famous painting The Third of May.

An application: counting double

If there is one thing that characterises the world of combinatorics it is that it is common for very simple ideas to give rise to incredible and unexpected results. That is the case of what is known as the *technique of double counting*. Basically, the purpose behind this technique is to count the same set in two different ways. With this principle we can solve a curious question about maps.

Let's begin by explaining the method with an example. Imagine that we want to hold a dinner party at our home with our closest friends: Mary, Peter, John and Anne. We ask each of them to bring some food and drink to the dinner. As we are very well-organised people, we have prepared a table showing what things each of our friends should bring:

	Cheese	Ham	Eggs	Beer
Me		X	X	X
Mary	X			X
Peter		X	X	
John	X	X		X
Anne		X	X	

So, for example, Mary will bring cheese and beer, while Peter will bring ham and eggs. The important point to note about this list is that it can be read in two different ways. On the one hand, if we read it horizontally, Mary will see what products she must bring to the party (cheese and beer) and, on the other hand, if we read it vertically, we shall see those charged with bringing cheese (Mary and John). The crux of the matter is that we can count in two different ways in order to find out *how many* items will be brought in total, Either by looking at those which each person must bring, or by looking at which people will bring each of the items.

Let's generalise the problem so as to be able to use the principle of double counting. Suppose a subset of the Cartesian product $C = A \times B$ and remember that the Cartesian product of A and B is the set of ordered pairs (a,b) in which it holds that a and b belong to the sets A and B respectively. We shall consider only some of the elements of the Cartesian product. In our previous example, the element (Mary, cheese) is valid, while the element (Mary, ham) is not, as Mary brings cheese but not ham.

To abstract this question, let's write $A = \{a_1, a_2,..., a_r\}$ and $B = \{b_1, b_2,..., b_s\}$. Let's now imagine that we have a generic table similar to the one that we designed for organising the dinner party, except that in this case the sets are abstract:

	b_1	b_2	b_3	b_4	...	b_s
a_1	X		X	X		X
a_2		X				
a_3		X	X			X
a_4				X		
a_5	X					X
$\vdots$					$\ddots$	$\vdots$
a_r		X	X		...	X

In the diagram we draw a cross to show that we are taking into account the pair under consideration, while we do not draw a cross if the pair under consideration does not form part of our set. The problem arises from knowing how to count the total number of crosses in a table like the previous one. The philosophy of the double counting method is that this count can be calculated in two different ways and that the same count will be obtained in both cases.

The first way of adding is counting by rows: we fix an element a_i (where the subindex i varies between 1 and r) and look at how many crosses appear in its row. If we carry out this operation for each of the possible subindexes and then add, the result will be the total number of crosses. Let's look at another way of obtaining the same count. Instead of adding in rows, we shall do it in columns. In other words, for each element b_j (where now the subindex j moves between 1 and s) we look at how many crosses there are in the associated column, and finally we add up all the columns.

This other way of looking at the same sum is called the *double counting method*. Notationally, using the hypotheses above, the double counting method is condensed into the following mathematical formula:

$$\sum_{a\in A}\left|\left\{(a,b): b \in B\right\}\right| = \sum_{b\in B}\left|\left\{(a,b): a \in A\right\}\right|.$$

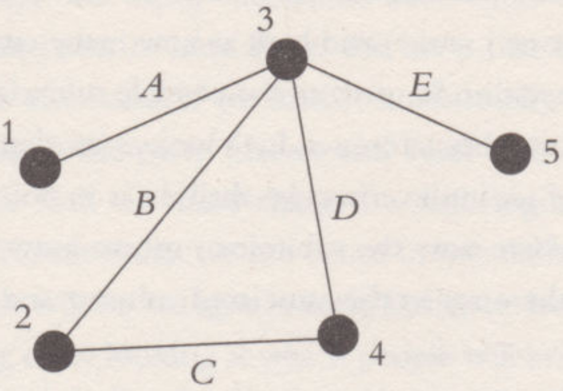
Let's look at how we should understand this formula: the symbol Σ (the Greek letter sigma) means 'sum'; the colon : is read as 'such that'; $\in$ stands for 'is an element of'; the vertical bar | is 'cardinal' or 'number of elements', and the curly brackets (or braces) {} denote a set. The first addend tells us that we are adding by rows (for each element of A we count how many there are in its corresponding row), while the second one indicates that the addition is carried out by column.

Let's look at an application of this method in the context of graph theory. To start off the reasoning, let's consider a graph G whose set of vertices and edges we shall denote by V and A, respectively. Let's now consider the following set:

$$C=\{(v,e):v\in V,\ e\in A,\ e \text{ is incident with } v\}.$$

Observe that this is a subset belonging to the Cartesian product $V\times A$, as for each vertex we consider only the edges that are incident with it. We shall apply the double counting technique to this set so as to deduce some sort of useful information. The key observation in the matter is this fact: each edge is incident *exactly* with two vertices, as an edge is defined by its two tips. In the following example, the set of pairs involved is:

$$(1,A),\ (3,A),\ (2,B),\ (3,B),\ (2,C),\ (4,C),\ (3,D),\ (4,D),\ (3,E),\ (5,E)$$



A graph with five vertices showing the set on which we shall apply the double-counting method.

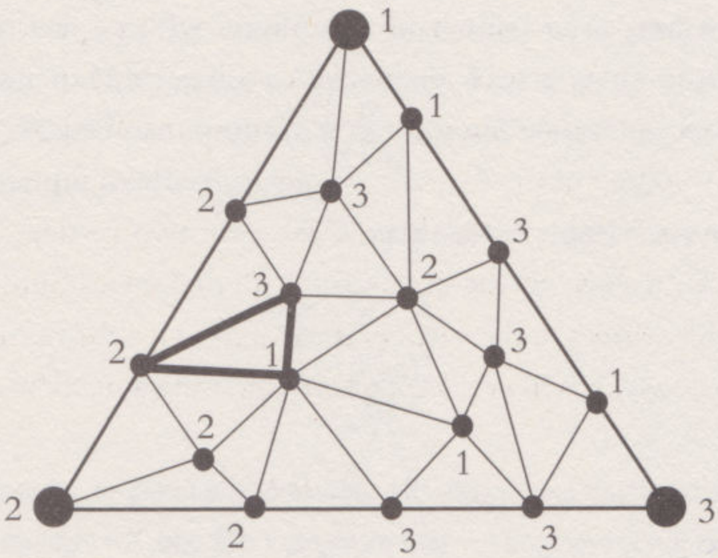
We shall introduce a little notation so as to make the mathematical reasoning that follows as easy as possible. Given a vertex v we say that the *degree* of v is the number of edges that are incident with it. We shall denote this value by $d(v)$. Now let's go on to apply the enumerative argument. The cardinal of the set C can be calculated in two different ways: on the one hand, by adding with respect to the edges (and taking into account what vertices are incident with each of them), or adding with respect to the vertices (and taking into account how many edges are incident with each of them). By adding in the first way we get twice the number of vertices, since, as we said before, each edge is exactly incident with two vertices. In the same way, an arbitrary vertex v is incident, by definition, with $d(v)$ edges. Summing up, we find that the sum of the degrees of the vertices is equal to twice the number of edges. The second term can be written in condensed form by using a mathematical summation:

$$2|A| = \sum_{v \in V} d(v),$$

where we specify that the degrees on all the possible vertices v are added. Note that this relation is valid for any graph and also for any map, as it only depends on its skeleton. Rather than being just a general formula without any clear application, from it we can actually extract very interesting structural information. It should be remembered that the sum of two even or odd numbers is an even number, while the sum of an odd and an even number always results in an odd number. Because the term on the left in the equation above is an even number, the result is that *the number of vertices with an odd degree must be even.* Indeed, if it were not so then the

sum on the right would be an odd number, which is impossible for reasons of parity. This observation now seems less obvious and, in fact, when it is subjected to a more in-depth study, the result grew in significance as *Sperner's lemma*, one of the true 'jewels in the crown' of combinatorics. It is, however, still elementary combinatorics.

Suppose that a triangle with vertices labelled with marks 1, 2 and 3 is subdivided into small triangles. That is to say, starting from the initial triangle we draw an additional set of vertices on the edges and in the interior. Then we draw additional edges in such a way that each face is incident with exactly three edges. Starting from this configuration, we go on to label the rest of the vertices by using the initial labels: 1, 2 and 3. The only condition we set is that the vertices that are on the edge determined by 1 and 2 can only be marked with the label of 1 or 2 (and in the same way for the triangle's other two edges). Sperner's theorem states that under these conditions there will always be a triangle with its three vertices having different labels. The following figure gives an example of this labelling; the thicker line shows a triangle that fulfils the desired property.



An example of Sperner's lemma. The thicker line shows the small triangle that fulfills the property.

The demonstration of this result is based on a very ingenious application of the double-counting principle. Proving the lemma requires that the number of good triangles (that is, those with vertices that have three different labels) must be other than zero. In actual fact, the number of triangles with this property is not revealed. Instead what is proved is that it must be an odd number. So, there will be at least

EMANUEL SPERNER, HIS THEOREMS AND THEIR CONSEQUENCES

There are many surprising applications of Sperner's lemma. One consequence of it was the *Brouwer fixed point theorem*, in which the existence of fixed points for certain functions was proven. What is extraordinary is that before it was known how to apply the Sperner lemma, all the proofs of this result that were known used methods that were far more sophisticated than the combinatorial arguments that we have outlined. That is why Sperner's lemma is seen as one of the first results of what is known as *combinatorial topology.*

Emanuel Sperner (1905–1980) carried out his research into mathematics at the University of Hamburg. His most important contributions were in the context of combinatorics, and more particularly in the result that we have shown here (commonly known as *Sperner's lemma*) and in the Sperner theorem on sets of a fixed size without mutual intersection: given the set $A=\{1, 2, ..., n\}$, a *Sperner family* is a family of subsets such that there is no pair of them in which one is contained in the other. For example, the family {{1}, {2}} is a Sperner family, while the family {{1}, {1,2}} is not, because {1} is a subset of {1,2}. Within this context, what is the maximum cardinal that a Sperner family can have? Sperner's theorem states that it cannot have more than

$$\binom{n}{n/2}$$

elements if n is an even number, or more than

$$\binom{n}{(n-1)/2}$$

if n is odd. These binomials are, in fact, the cardinal of the family of subsets of $\{1, 2, ..., n\}$ with $n/2$ elements if n is even and $(n-1)/2$ elements if n is odd. For instance, for that family it will be:

$$\{\{1,2\}, \{1,3\}, \{1,4\}, \{2,3\}, \{2,4\}, \{3,4\}\}.$$

This is one of the fundamental results of combinatorial lattice theory and of partially ordered sets, and a very particular case of a more general theorem called *Dilworth's theorem.*

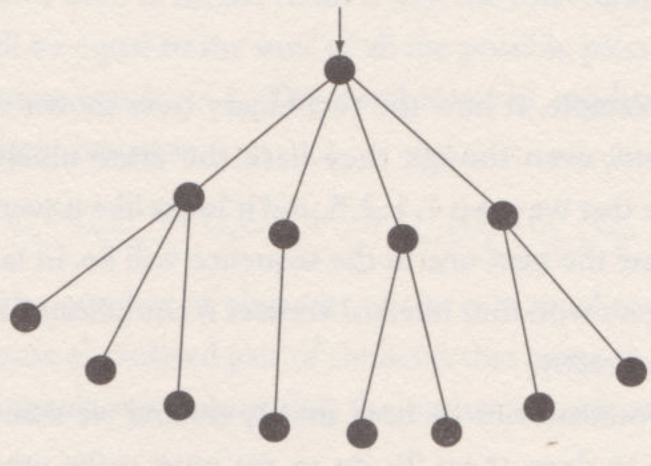
one triangle with that property! Any reader interested in taking an in-depth look at this result can find the details in the Appendix.

Trees: key players in graph theory

Let's return to maps and counting. Up to now we have shown that there is a great difference between graphs and maps, and that being able to draw a graph in such a way that the edges do not intersect does not depend on our skills but on the intrinsic conditions of the combinatorics object.

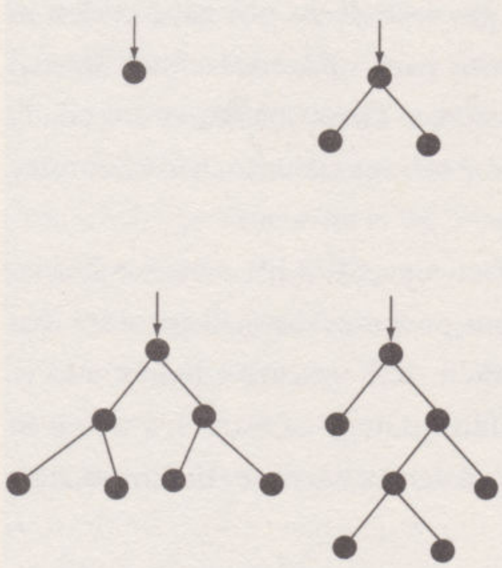
We shall now take a more in-depth look at graphs that do not have cycles, in other words, graphs without closed paths between pairs of vertices. Such special graphs are known in the world of mathematics as *trees*. These structures are vitally important in computation, data storage and in more remote fields such as chemistry and biology. Their name comes from their shape.

A tree drawn on the plane defines only one face, since, as it has no cycle it does not have an *internal face* and an *external face*. For our purposes we shall consider that the trees have a *root,* i.e., the vertex that is marked with an arrow below, and is, therefore, privileged with respect to the others. This strategy of rooting a tree is in fact very much allowed and means that the marked vertex becomes the initial state of the system that the tree is modelling.



A tree with a root. The map only has one face (the external) and has no cycles.

Let's look at one extremely important subfamily of trees – binary trees. They are the ones with vertices that are incident with either three edges or with just one edge. In the first case we say that the vertices are *internal*, while a vertex of degree one is called a *leaf* of the tree. The vertex that holds the root is also considered to be an internal vertex, as the root makes the privileged vertex degree increase by one unit. With these hypotheses, how many binary trees can be drawn with a fixed number of internal vertices? Let's look at the first binary trees:

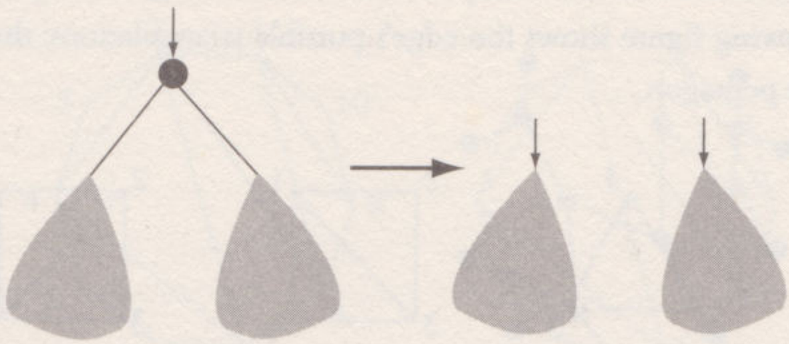


Binary trees with 0, 1, 2 and 3 internal vertices.

Let's look, for example, at how the two binary trees shown with three internal vertices are different, even though they have the same underlying graph. The numerical sequence that we get is 1, 1, 2, 5, and it looks like it would be complicated to deduce what value the next one in the sequence will be. In fact, discovering the number of binary trees with four internal vertices is complicated if we want to carry out the calculation directly.

By using the combinatorics we have already studied we shall be able to deduce it without needing to draw them. To do so, we must make use of the recurrence structure of the trees. See how the objects that are joined to the root vertex are again trees, with a determined number of internal vertices. This idea is the one that is shown in the next figure, in which we decompose a binary tree into two smaller

ones. The inverse operation is also simple and consists of attaching the two roots to a new vertex.



Decomposition of a binary tree into two smaller binary trees. The total number of internal vertices is reduced by one unit.

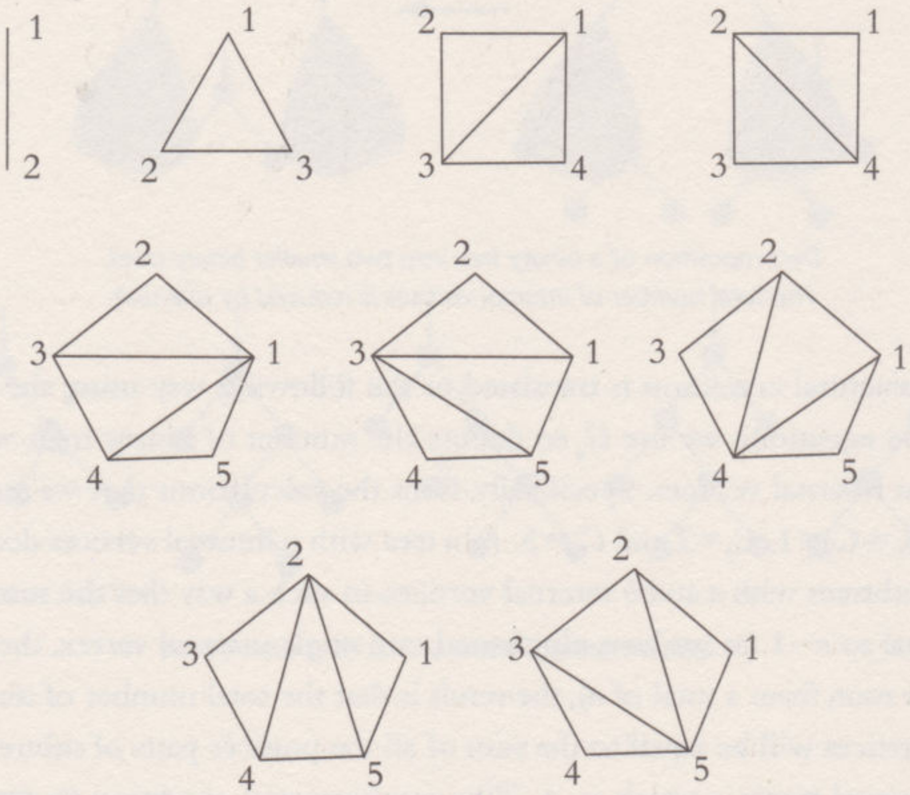
This structural condition is translated in the following way using the language of recursive equations: we use C_n to denote the number of binary trees with roots and with n internal vertices. Specifically, from the calculations that we have made, we have $C_0 = C_1 = 1$, $C_2 = 2$ and $C_3 = 5$. As a tree with n internal vertices decomposes into two subtrees with a and b internal vertices, in such a way that the sum of a and of b is equal to $n-1$ (as we have eliminated one single internal vertex, the one that carries the root, from a total of n), the result is that the total number of trees with n internal vertices will be equal to the sum of all the possible pairs of subtrees whose sum of internal vertices equals $n-1$. This combinatorial condition transposes into the following recurrent equation:

$$C_n = C_0 C_{n-1} + C_1 C_{n-2} + \ldots + C_{n-2} C_1 + C_{n-1} C_0,$$

which tells us that the number of elements of size n is equal to the sum of all the possible ways to choose an ordered pair of elements that come to $n-1$. By substituting values in this formula we deduce that the numerical sequence for the binary trees is as follows:

1, 1, 2, 5, 14, 42, 132, 429, 1,430, 4,862, 16,796, 58,786, 208,012, 742,900, 2,674,440,...

What is really curious in this sequence is not its formula but rather that there are other combinatorial objects that are also counted by this same series. Let's define another family of objects that on the face of it has nothing to do with binary trees. Let's take a regular polygon of n sides, with the vertices labelled 1, 2, ..., n anti-clockwise. Let's look at the number of decompositions into triangles of the interior of the polygon. Each of these decompositions is called *triangulation of a polygon with n sides.* The following figure shows the edge's possible triangulations: the triangle, the square and the pentagon.

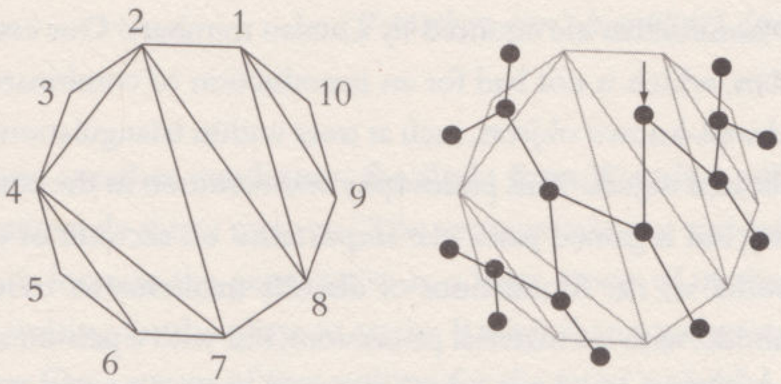


Triangulations of polygons with three, four and five sides.

It is curious that in this family the same numerical sequence is repeated: 1, 1, 2, 5. Could it be that those families, even though they are different, are counted by the same numerical sequence? Yes, they are. Let's see why.

To prove that for each triangulation we can associate it with one single binary tree we shall use the *dual map* of a given map. This map is built in the following way from an initial map. Its vertices are drawn in the interior of each of the faces of the primitive map and two vertices are joined by an edge if the faces with which they are associated are incident. This construction is transposed into the triangulations

in the following way: we draw a vertex on each of the faces of our triangulation. Additionally, we draw a vertex of degree one (i.e. a leaf) for each edge of the polygon that is not defined by vertices 1 and 2. To this latter edge we shall associate the root of the binary tree. Look at the example in the following figure, which shows the construction for the case of a decagon.



Triangulation of a 10-sided polygon and the associated dual construction.

Note that the dual map contains no cycles. If it did, the triangulation would have to have an interior point, and that is not possible by definition. Furthermore, that map only has vertices of degree three (those associated with the triangles) and of degree one (those associated with the edges of the polygon). Therefore the map is, in fact, a binary tree, with internal vertices that are those associated with the triangles of the triangulation. Reciprocally, we can show that for each binary tree with n internal vertices we can construct a triangulation of a polygon of $n+2$ sides, the dual map of which is precisely the initial binary tree. In this way there is a bijective relationship between triangulations and trees and, therefore, of a certain size there are as many as of the others. The size is, incidentally, the number of vertices of the polygon and the number of internal vertices, respectively.

The reader may find it fascinating that the equation given above, so magical and mysterious, should appear in different enumerative problems. The solution to this recursive equation is well known and gives rise to what are known as *Catalan numbers*, named after their discoverer, the Belgian mathematician Eugène Catalan, and whose expression is as follows:

$$C_n = \frac{1}{n+1}\binom{2n}{n}.$$

Over and above combinations and binomial numbers, Catalan numbers show up in enumerative combinatorics as the main expressions that cannot be obtained by means of direct counting. In fact, Catalan numbers appear in a wide variety of problems and contexts. It is for that reason that Richard Stanley, one of the most important researchers in the field of enumerative combinatorics, in problem 6.19 of his reference book *Enumerative Combinatorics* shows more than 60 different combinatorics families that are counted by Catalan numbers! Our contribution has been two families, which is not bad for an introduction to combinatorics.

We can find well-known objects, such as trees within triangulations, even in very midst of complicated objects. This philosophy was common in the combinatorics of the 20th century, but it gained particular importance on account of one important person who would set the foundations of discrete mathematics trembling. A man with no fixed abode, with no material possessions, but with a passion and dedication to mathematics as had never been seen before. An eternal nomad with a desire to widen the appeal of mathematics and the great enigmas of knowledge.

LOST IN A DICTIONARY OF NUMERICAL SEQUENCES

In the same way that writers use a dictionary of synonyms and antonyms to enrich their text, mathematicians also need resources to help them in their research. Let's imagine that we are counting a certain family but we want to find the general term. No problem. If there is anyone who has studied the same sequence we can find out by looking it up in *The On-Line Encyclopedia of Integers Sequences*, on the web site http://oeis.org/. This site is maintained thanks to the work of the researcher Neil Sloane, of the AT&T company's research laboratories. Each sequence discovered is noted down and catalogued. In this way, when researchers need to know if the sequence they have found has appeared previously, or if it also counts something else, they only need to key the first terms into the search engine and see what appears.

Just out of curiosity, the reader can try it out by keying in television's most famous and enigmatic numerical sequence, i.e. the sequence from the American TV series *Lost*: 4, 8, 15, 16, 23, 42. He or she will see that the sequence is catalogued, and that it even has a reference number in this numerical encyclopaedia: A104101.

Chapter 3

The Eternal Nomad

If numbers aren't beautiful, I don't know what is.

Paul Erdős

"Due to adverse weather conditions, the flight from Waterloo will arrive with a delay of approximately thirty minutes." The message over the airport's loudspeakers simply serves to increase the expectation in a large group of mathematicians who are nervously waiting for the plane to arrive. Rather than a gathering of academics, they look more like a group of fans waiting for the latest rock idol to appear. This excitement is well justified, however, as it's not every day that a living legend appears in person.

After a longer than usual delay the airliner lands and within a couple of minutes the passengers begin to emerge. One of them stands out from the rest: elderly, grey-haired, fragile and looking somewhat dreamy and hesitant. He wears an old, dark-coloured jacket and in one hand carries a little cloth suitcase, while the other holds a fistful of crumpled papers. Many of his fellow passengers might think that the ragged old man is a tramp, even mad; few of them would guess that they have just shared their flight with one of the most lucid brains of our times.

The fellow's appearance caused a stir amongst the group of erudite mathematicians awaiting him. Eager faces, expectation, unchained nerves all around; everyone is excited. The old man shakes off his look of seemingly imperturbable, tantric concentration. His lips shaping a smile, he looks round at this horde of thinkers and says: "My brain is open!" It is the beginning of another of Uncle Paul's visits, bringing with it new and exciting mathematical challenges, unsolved puzzles and theorems to be discovered.

My brain is open!

In all arts and sciences there have always been and always will be figures who define a turning point in their fields. Paul Erdős was one of those people. In the second

half of the 20th century the scene described above was repeated on numerous occasions all over the world. The reason behind the travels was always the same – mathematics and the intense passion that Paul Erdős held for them. His work in mathematics could be described in different ways depending on who did the describing. Leaving aside judgements on scientific matters, the most outstanding aspect of his legacy is its extension and variety; his work is far wider-reaching than that of any other mathematician. If we were to classify all mathematicians throughout history according to the number of written works they have produced, Leonhard Euler would win hands down. The great scientist from Basle is, up to the present day, the mathematician to produce the largest number of written pages. Euler would be the champion by weight, but Erdős would win as regards number of articles published. A total of more than 1,500 articles are testimony to his scientific output. In these works, Erdős nosed around in a wide variety of fields in pure science, from mathematical analysis to geometry and on to algebra. However, despite this huge mathematical output, Erdős is particularly renowned for his devotion to a type of questions of a discrete nature. It was in this world, in the field of finite geometry, of graphs, of shrewd counting on the fingers and of additive properties of integers, where our protagonist made his key contributions.

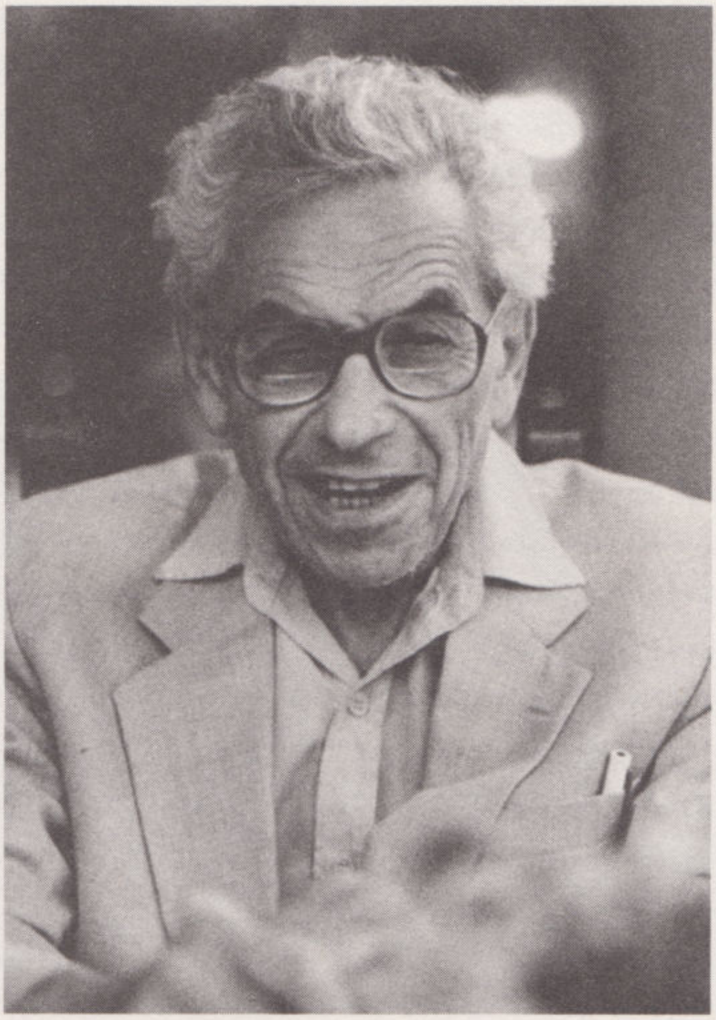
Apart from the works he produced, Paul Erdős was also instrumental in combinatorics first being accepted as a discipline in its own right in the Mathematics Olympiad. Prior to his work, problems of a discrete nature were considered to be specific issues or even simply anecdotes and ingenious games which were subordinate to their big brothers in the science hierarchy. Together with the problems of combinatorics and the branches that developed from them, Erdős's contributions gave substance to the notion of combinatorics as we understand it today, with its own problems and techniques.

Leaving aside this genius's contributions to science (which is not easy to do), this scientific legend managed to turn mathematics into a social activity, if by that we understand a creative activity in which more than one person takes part. If we were to carry out a study of the bibliography of science works prior to the second quarter of the 20th century, we would spot a general pattern: most of them had just one author. Erdős attempted to break this unwritten rule by creating a wide network of collaborators. This philosophy led to him being able to count on over 500 collaborators... that's more than the number of acquaintances that most people have! This network of contacts was consolidated thanks to the nomadic existence

he led throughout the greater part of his life; a life full of journeys from one university to another, from institute to institute, continent to continent, always accompanied by his (very few) material possessions.

There is ample justification for Paul Erdős to act as our *cicerone* along this pathway that will lead us into the exciting world of modern combinatorics, a discipline in which Uncle Paul (as he liked to be called) made vital contributions by forming puzzles, by conjecture and by investigation. Many of these enigmas have been solved, but some have not, and not because the most privileged minds of the 20th century have not tried.

But before setting out on our expedition to these enigmas, we shall take a better look at the full magnitude of the scientist, man and legend, since it is far better first to get to know one's guide so as to get more and better enjoyment from the journey.



Paul Erdős, a figure who marked a turning point in the world of combinatorics.

Childhood

The intellectual, artistic and scientific bustle in Budapest's inter-war period was intense. Instead of being a natural frontier between the cities of Buda and Pest, the Danube was more like a link uniting different communities into an outstanding melting pot of cultures. In the cafés, boulevards and parks of the Hungarian capital the atmosphere of artistic creation was comparable to that of Paris and London. That cultural explosion was due in great part to the local Jewish community which had coexisted in liberty and complete harmony with the city's other cultures for decades. The political situation in the country afforded Jews full rights, and allowed them to hold public office, which were rights established back in the second half of the 19th century during the dual monarchy of Austria and Hungary. Such was their economic influence and cultural contribution in the Budapest of the time that some people took to contemptuously referring to the city as 'Jewapest' instead of Budapest.

This anti-semitic undercurrent would a few years later lead to the terrible confinement of the Jewish community to the city's ghetto, and their later displacement to horrendous death camps. But all that barbarity for humankind was still to come and no one could have foreseen such events during the period between the Great War and the next global conflict to come.

It was in that historical context and in that city of Budapest that Paul Erdős, born to Jewish mathematicians Anna and Lajos Erdős in 1913, was to grow up. Paul never got to know his elder sisters, Magda and Klára, as they lost their lives in an outbreak of scarlet fever that blighted the city. Following that family tragedy, the young couple of mathematicians focused all their energy on giving Paul a good education, love and affection, and even perhaps being too protective of him. Times became difficult for the young couple in a Budapest where anti-Semitism was chillingly beginning to emerge. However, that was not the couple's only problem: just a year after Paul's birth the Great War broke out, sparked off by the assassination of Franz Ferdinand of Austria in Sarajevo. In reaction to that provocation, and as a symptom of the extraordinarily tense political situation, the Austro-Hungarian Empire declared war on Serbia. Paul's father, Lajos, was called up.

It was not long before tragedy struck the family: Russian troops captured Lajos and sent him as a prisoner to feared, frozen Siberia for six long years. Uncertain of whether she would ever see her husband again, Anna dedicated all her love and efforts to raising her little boy, with the result that the young Paul was prevented from

A SOCIETY ENTHUSED WITH INTELLECTUAL FERVOUR

If we were to make a list of all the intellectuals who grew up in the Budapest of the interwar period, it should come as no surprise that we would be noting down the names of many of the greatest scientific and artistic minds of the 20th century. In mathematics, besides Paul Erdős, many other young people (and not so young ones) would become the authors of new theories and discoverers of new pathways. Such was the case of Leopold Fejér, who years later was to be the doctoral thesis advisor for Paul Erdős and one of the progenitors of modern mathematical analysis. Under his protection, he gathered together a school of young, brilliant mathematicians in which Paul Turán, George Pólya and Tibor Radó were just three of those who stood out. Another of his famous students who should be mentioned is John von Neumann, who was to lay the formal foundations of present-day quantum physics as well as becoming one of the fathers of computational game theory. Years later, now in the United States, he would be one of the first civilian members of the Manhattan Project, set up by the US government to develop the atomic bomb.

The Hungarian legacy to humankind is not limited solely to mathematics: the aeronautical engineer Theodor von Kármán was a pioneer of supersonic flight research, and the Nobel Prize winner George Hevesy provided new applications for radioactivity in numerous areas. And, naturally, mention must be made of the country's artistic legacy, beginning with music (Georg Solti, Béla Bartók and others), painting (László Moholy-Nagy), the cinema (Alexander Korda) and showbusiness (Erik Weisz, known all over the world as Harry Houdini).

A scene from Budapest in 1910.

going to school and educated at home by his mother and a private tutor. During those years, a truly intense relationship became established between Paul and his mother. Years later, Anna would become Paul's tireless companion in his seemingly endless nomadic existence.

Paul's talent did not go unnoticed by his mother. It is said he learnt to count before he could walk. By the age of three he was able to do additions, and by the time he was four he could do complicated calculations such as long multiplication - even working out the number of minutes that a person had lived. Discovering negative numbers while still a young child opened this genius's mind more than ever and showed him that in mathematics there are no limits to the intellect. Paul Erdős's gifts were favoured by an environment that was very propitious for intellectual pursuits. The Hungary in which he was growing up had an education system that was very solid and based on the detection and furthering of individual qualities. Primary and secondary school teachers were well appreciated for their work and took their responsibilities further than simply imparting lessons; they took very seriously their young pupils' upbringing and education. The result was that it was common for prestigious researchers to dedicate their time to discussing problems with talented pupils.

One of the numerous ways in which that inter-generational dialogue was achieved was through magazines dedicated to mathematical problems, such as the popular *KöMaL* journal of the time, a publication founded in 1894 by Professor Daniel Arany, and which, in his own words, was "...to give a wealth of examples to students and teachers...". (The journal is still in business and can be consulted on its web site: *www.komal.hu*.) Secondary school pupils were invited to solve problems that appeared in a monthly bulletin, which was a tool for detecting innate talent for sciences. But it was not child's play, as the researchers themselves also devoted their time to thinking about problems of this type and to solving them. The reason is that, when all is said and done, mathematicians devote themselves to this discipline for the excitement and for the curiosity they have for finding solutions to enigmas.

Years later, this philosophy would lead to the creation of maths competitions at a local and national level in Hungary, which were the embryo of what is nowadays the International Mathematical Olympiad (and also the olympiads for physics, chemistry and even information technology), which are competitions at the highest level for secondary school pupils. In the same way that sports Olympics require a lot of

preparation, these scientific competitions require a series of selection phases, a lot of work, and a great deal of ingenuity too. It is, in fact, in these competitions where many of the greatest scientists of the second half of the 20th century began their careers in mathematics.

But let's get back to Budapest and the life of the young Erdős. The situation in Hungary simply got worse with the end of World War I. Antisemitism and anti-communism became altogether more potent in 1920 when Miklós Horthy Nagybánya, a right-wing nationalist and commander of the Austro-Hungarian army, took control of the country. With Horthy in power, the Jewish community found itself subject to new laws that were similar to those Hitler was to introduce in Germany 13 years later and which would lead to the barbaric situation of Nazi Germany. Not everything was bad, as Lajos returned home after his term as a prisoner of war in Siberia. Within that historical context, the curiosity and potential of the young Erdős simply increased, no doubt stimulated by the city's cultural tradition. It was in Budapest that as a teenager he met some of those who would be his finest companions in his travels throughout the world of mathematics, such as Paul Turan, George Szkeres and Esther Klein, all of them enthusiastic problem solvers for student magazines. This friendship would lead the small community of young mathematicians to make joint reflections that years later would be the cornerstones for the creation of new and beautiful theories.

And as time went by, in a difficult environment on account of the hatred that was steadily growing, Erdős grew up and began his career in mathematics, first as a student, then completing his thesis and finally leaving his beloved Budapest.

Teenage years and exile

The enthusiasm and capacity shown by Erdős for solving mathematical problems became increasingly evident to such an extent that his potential was recognised at an international level when he was still very young. One of the first important results he obtained, and at the age of only 19, was an alternative proof of what is known as *Bertrand's postulate*.

To understand this theorem it is necessary for us first to recall the concept of a *prime number.* We say that a number is prime if its only divisors are 1 and itself. So, for example, 2, 7 or 11 are prime numbers, while 6 is not, as it can be divided by 2. Bertrand's theorem states the following:

"Between a number and its double there is always a prime number."

For example, between 4 and 8 there is the prime number 5, and between 13 and 26 there is number 17, also a prime. The first proof of this result was produced in 1850 by Tchebychev by using arguments that were quite sophisticated. Years later, the self-taught Srinivanasa Ramanujan found a simpler proof, though it still included certain technical terms. Finally, in 1932 Paul Erdős conceived an elementary and extremely elegant proof by making use of binomials, which we have covered extensively already.

It was at this time that close cooperation was forged between Erdős and his Hungarian contemporaries. Like Erdős, many of these contemporaries had been involved in the maths journals of that era, which made it possible for talented youngsters to meet up later in their teenage years to take part in heated discussions on mathematical issues. They had a fixed meeting place: the statue of the unknown writer in the city park (Városliget) in Budapest, immediately in front of Vajdahunyad Castle. The sculpture there was erected in honour of the "anonymous notary of the glorious King Béla". That unidentified writer is acknowledged to be one of the first chroniclers of Hungarian history, from back in the 12th century. The local superstition is that touching the writer's pen brings good luck, and that's why the bronze pen is always so shiny. The unknown writer's pen was to bring Erdős and his companions luck on the pathway to discovering knowledge.

Erdős's university career was short and fast. University entrance restrictions imposed on the Jews were not a problem for him since he was assured a place there as the winner of the national entrance exams. And so, in 1930 he entered the Pázmány Péter University in Budapest. He was to achieve his doctorate there in 1934, under the tutorship of the great Hungarian mathematician Leopold Fejér. Here again his genius stood out from the rest, as he was awarded his doctorate when he still only 21. (Note that he had entered the university at 17, and therefore obtained his degree and doctorate in four years flat!)

The situation for the Jewish people grew more serious, and the community came to fear the worst, which caused Erdős to decide to leave. The situation of a Jewish scientist was, to say the least, difficult. His destination was England, and more specifically, Manchester University. This was thanks to an invitation from the great Jewish-American mathematician, Louis Mordell. There is an anecdote that Erdős himself used to tell about his arrival in Manchester, a place he knew nothing about:

The statue of the unknown writer, known as the Anonymous statue, the meeting place of Erdős and his mathematical colleagues.

"I arrived mid-afternoon at Mordell's house not having eaten anything all day. At five o'clock they served tea, and I was starving. I felt so ashamed at the thought of having to confess that I had never spread butter on a piece of toast that I decided to imitate what the others were doing, and discovered that it was not such a difficult task as it appeared..."

That little story clearly shows the excessive pampering that Erdős had been subject to by his mother during his childhood and adolescence. This was the first time in his life that he had had to look after himself, and so he did. From that first visit to England onwards, Erdős began to travel assiduously, a habit which years later would become

his way of life. He met the great researcher Godfrey Harold Hardy at Cambridge that same year of 1934, and Stanislaw Ulam a year later. Years on, that second meeting was to be key to opening the doors to New York for him.

While Erdős was travelling around all over Britain, the situation in Europe was becoming critical. This, however, did not prevent Erdős from returning to his beloved Budapest three times a year during the period he lived in Manchester. However, the situation deteriorated further, and in 1938 Hitler took control of Austria. On account of the upsurge in Nazi activities (and, in particular, the events of 3 September in the Sudetenland, in the former Czechoslovakia), Erdős made up his mind to leave Europe and emigrate to the United States where he would settle into the Institute for Advanced Study at Princeton, an institution of the highest intellectual and scientific level. Erdős left Budapest for the last time, leaving behind his parents and friends, and not knowing when he would next be able to visit. He would, in fact, not see his parents again for 10 years. On 18 September, leaving London and bound for New York, Erdős waved goodbye to the Old World... for the time being.

A RESEARCH INSTITUTE OF THE HIGHEST LEVEL

The Princeton Institute for Advanced Study was founded in 1930 with the aim of providing the leading researchers in their own particular field with the best work environment possible, with no need for them to spend their time teaching. While there are informal links and a great deal of cooperation goes on between the two, the Institute is an independent organisation and does not form part of Princeton University.

Its *modus operandi* is very different from what is normally understood for a university. It keeps a small number of professors who are complemented by visiting members selected annually, and the researchers are given complete freedom to develop their own scientific projects. The research is not carried out through contracts or subject to external control but in accordance with each scientist's own criteria, with funding being provided through donations and subsidies, and attendees not being charged any fees. As the statutes from when it was founded show, among its objectives was that of taking in Jewish researchers who were not allowed to aspire to research positions in Princeton due to its institutionalised antisemitism. That is the reason why great researchers of the 20th century such as Albert Einstein, Robert Oppenheimer ('the father of the atomic bomb') and John von Neumann held positions at this institution. Kurt Gödel, though he was not Jewish, did his thesis with Hans

The United States, Israel… Life as a nomad

Erdős's life in the United States began at the Institute for Advanced Study at Princeton. However, his stay there was shorter than expected: after a year at the prestigious institution they did not renew his scholarship. The official reasons given were that there were not enough funds, but the rumours were that the real reason was that the spirit of cooperation that Erdős displayed was upsetting the great thinkers installed in the institute at that time. In any case, Oswald Veblen, at that time director of the institute's department of mathematics, managed to get Erdős funding of $750 per month, which would be enough to last him for the coming academic year.

After moving to the University of Pennsylvania in 1941, an incident seemingly of little importance took place on Long Island: Erdős and another two researchers were having a heated discussion when, totally unjustifiably, they were arrested by the police. What the absent-minded mathematicians had not noticed was that during their discussion they had strayed over a certain line forbidding public access, which resulted in Erdős having a record on file with the FBI. Such a

Hahn, who was, and was also one of the illustrious members of the institute during its first years of existence. In spite of its idyllic environment for research, there were many eminent voices who criticised it. Such was the case of the controversial and almost mythical physicist Richard Feynman, who, in his well-known book *Surely You're Joking, Mr. Feynman*? says this:

"... When I was at Princeton in the 1940s I could see what happened to those great minds at the Institute for Advanced Study, who had been specially selected for their tremendous brains and were now given this opportunity to sit in this lovely house by the woods there, with no classes to teach, with no obligations whatsoever. These poor bastards could now sit and think clearly all by themselves, OK? So they don't get any ideas for a while: they have every opportunity to do something, and they're not getting any ideas. I believe that in a situation like this a kind of guilt or depression worms inside of you, and you begin to worry about not getting any ideas. And nothing happens. Still no ideas come. Nothing happens because there's not enough real activity and challenge: you're not in contact with the experimental guys. You don't have to think how to answer questions from the students. Nothing!...".

banal event as this was, during the McCarthy era, to cause Erdős problems with the authorities.

Erdős's mind was in the world of maths, but his heart was on the other side of the Atlantic. In 1943 the situation of his family and country was a great worry to him: He had had no news of his family since 1941, and would get none until the city was liberated in 1945. The news that his father had died some years before, in 1942, was a great blow to him. And that was not all: the Nazi reprisals on the Jews had been particularly hard on the community in Budapest and many of them had died in the Auschwitz concentration camp. Erdős's family was no exception, and several of his relatives died in the Holocaust. Anna, his dear mother, had survived. It was not till the end of 1948 that Erdős was able to return to his country to visit his loved ones and to rebuild the links broken during his long stay in the United States. After that year he resumed his journeys between the Old World and the New World and became a temporary lecturer-researcher at the Notre Dame University in Indiana.

The year 1954 was a turning point for our protagonist. The International Congress of Mathematicians was to be held in Amsterdam, and Erdős was invited as a speaker. As a foreigner, he had to request a visa to return to the United States, but his extensive correspondence with numerous mathematicians, including some behind the Iron Curtain, raised the suspicions of the immigration officials. On top of this, Erdős already had a record with the FBI due to his slip-up on Long Island. The suspicions were strengthened by the 'witch hunt' of communists during the McCarthy era. When he returned from his journey, the immigration officials did not allow him entry. Erdős recalled that "...the officials asked me all kinds of silly questions...". Things only got worse when the immigration officers asked him what he thought of Karl Marx, to which he replied: "...I'm not competent to judge, but there's no doubt he was a great man..." All in all, it led to him leaving the United States and to being condemned never to return to the 'land of opportunity' until 1958 when he was given a special visa to attend a conference. In relation to this business, Erdős used to say: "...The United States' foreign policy consists of two issues – not admitting Red China to the UN, and not admitting Paul Erdős to the United States..." The consequence of this was that Erdős emigrated to Israel and their Technion Institute, where he remained for over ten years. It was during this period that he truly began to develop his nomad spirit, which had shown itself both in his days in England and in the United States.

His relationship with his mother was once again strengthened on Erdős's return from North America. When he was not travelling he stayed with her, and she took care of him and every little detail of his life so that he could concentrate solely on thinking and creating mathematics. They had such a strong relationship that from 1964 onwards Anna began to travel with him though she was by now 84. His mother was his manager, his mentor and his advisor.

After Anna's death in 1971, Erdős' eccentricity became more and more evident. He refused to enter his parents' old apartment in Budapest and instead asked his friends to put him up whenever he was in town. He might turn up in the middle of the night, whatever time it was, and announce that he was ready to discuss mathematics with them. He began to work at a frenetic pace, around 19 hours per day aided by strong cups of coffee and amphetamines.

Paul Erdős with his mother.

The typical ailments of his increasing age (and drug use) started to take their toll but that did not prevent him from continuing with his hectic schedule. He refused an operation on his cataracts because it would have prevented him from working for a time. His heart was weak and he needed a pacemaker, which is not a complicated operation but does require hospitalisation. Erdős refused to spend the night in the hospital as it would have meant him missing part of the talks at the congress he was attending. The operation finally went ahead, and he and his two cardiologists went to the maths conference together.

He continued at this frenetic pace without rest for what was left of his life. By now, with his powers depleted, and lacking the mental agility of his youth, he carried on travelling and staying, like a nomad, wherever he could, discussing and proposing new enigmas. In 1996, during a conference in Warsaw, and at the age of 83, his heart stopped, leaving the great family of combinatorics researchers fatherless.

The personality of the genius: conjecture and proof

The unconventional path that Erdős took was to forge his character. It can be said that his basic interest in life was mathematics. As a person, and due to his great passion for the abstract world, perhaps the most notable factor was his disinterest in the material world: money, status and honours. He never worried about having a home or a position, and used the money from prizes he was awarded in helping young people with talent. Here is one example: in 1984 he won the Wolf Prize, one of the most prestigious in mathematics and of an intellectual level similar to that of the new Abel Prize.

This distinction brought with it a cash reward worth $50,000. Of this money, Erdős kept only $720 for himself. He donated the majority of the prize to the Technion Institute as a token of gratitude for having taken him in when the US authorities denied him entry to the USA. Part of that amount was used to fund a postdoctoral position in honour of his parents. He gave the rest to friends, relatives, students and colleagues.

Of the money he earned he offered a large proportion as prizes for people who managed to solve problems that he had not been able to work out. These prizes went from $1 (for problems that in his opinion were simple) up to $10,000 (for those of extreme difficulty and for which he did not foresee any significant progress

FAMOUS SAYINGS THAT DEFINE A CHARACTER

Paul Erdős's philosophy on life can be summed up in his saying "*conjecture and proof*", but beyond mathematics, Erdős's interests also covered politics, philosophy and theology. The quotes below, in which a little more of his character shows through, are good examples:

"You don't have to believe in God, but you should believe in The Book" (referring to the book of proofs).

"There'll be plenty of time to rest in the grave."

"God is the supreme fascist."

"Television is something the Russians invented to destroy American education."

"Problems worthy of attack prove their worth by fighting back."

"God may not play dice with the Universe, but something strange is going on with the prime numbers."

"I hope we'll be able to solve these problems before we leave."

"Some French socialist said that private property was theft ... I say that private property is a nuisance."

"It will be millions of years before we'll have any understanding, and even then it won't be a complete understanding, because we're up against the infinite."

being made in the coming years). These prizes were no more than excuses for folk to get working on solving one of Uncle Paul's difficult puzzles, as they were quite often paid with blank cheques. When the financial situation began to grow worse, his adviser and friend Ronald Graham (who took over the role of his mother Anna as soon as she died) began to pay his fees into one single account. It is interesting to note that those who managed to solve a problem wanted the prize... but only to frame the cheque bearing the living legend's signature!

There is, however, an immense number of his conjectures for which the solution is still not known and, in fact, for which researchers are nowhere near being able to provide an answer. One of them concerns the following problem, which Erdős thought up together with his friend from his teenage years Paul Turan, and for which he offered one of his largest rewards, $3,000.

"Let $A = \{a_1, a_2, a_3, ...\}$ be an infinite set of natural numbers. If the sum

$$\frac{1}{a_1}+\frac{1}{a_2}+\frac{1}{a_3}+\ldots$$

is divergent (i.e., infinite), then A contains arbitrarily long arithmetic progressions.

To make the conjecture comprehensible, we need to clarify some points. There are a number of concepts that the reader may find confusing. Firstly, what is an infinite sum? And what does divergent mean? It is obvious that if we have a finite number of numbers we can add them up (the first to the second, the result we get added to the third, and so on), and their sum will always be another number. However, if we have an infinite set of numbers we cannot proceed in the same way, as we would have to do an infinite number of sums, which is not feasible. By infinite sum we understand the *limit* of this procedure. We begin to add the first to the second, its result to the third, and so on. If these sums begin to get closer and closer to a certain number, we shall say that this is the sum of the series.

These infinite sums can have the following curious property: if, as we go on adding terms, the total sum grows uncontrollably (which in maths is known as *unbounded*), then it is said that the infinite sum is *divergent*. For example, the sum of the inverses of the natural numbers (also known as the *harmonic sum*) satisfies this property. We just have to take a calculator and begin to add to see that the result is the following sums:

$$\frac{1}{1}+\frac{1}{2}+\frac{1}{3}+\ldots+\frac{1}{9}+\frac{1}{10}=2.928968254;$$

$$\frac{1}{1}+\frac{1}{2}+\frac{1}{3}+\ldots+\frac{1}{99}+\frac{1}{100}=5.187377518;$$

$$\frac{1}{1}+\frac{1}{2}+\frac{1}{3}+\ldots+\frac{1}{999}+\frac{1}{1000}=7.485470861.$$

The more terms we consider, the greater the sum, and this sum becomes as big as we want.

The other point that needs to be clarified is the arithmetic progressions. We say that a set of n numbers is an arithmetic progression of length $n+1$ and ratio k if they can all be written as $a+bk$ where k takes values between 0 and n. The Erdős–Turán

conjecture tells us, therefore, that under weak conditions (we know nothing about the set, only that the infinite sum diverges) the set must have some well-defined sub-structures (the arithmetic progressions of an arbitrarily large length).

As we will see later, this innocent question, which was formulated by an old man, has turned out to be extremely profound and was the drive behind the development of some of the most important and complex mathematical theories of the second half of the 20th century. But let's not get too far ahead.

The philosophy of *conjecture* and *proof* led to a tireless search by collaborators throughout the world, regardless of ethnicity, nationality or politics. That enthusiasm for union and joint creation resulted in Erdős becoming the most prolific

A PROOF OF DIVERGENCE

Let's take a more in-depth look at how we can prove that the harmonic sum diverges. The first point that has to be taken into consideration is that if $a<b$, then

$$\frac{1}{b}<\frac{1}{a}.$$

From this, what we can see now is that the following inequalities are fulfilled:

$$\frac{1}{3}>\frac{1}{4};\quad \frac{1}{5},\frac{1}{6},\frac{1}{7}>\frac{1}{8};\quad \frac{1}{9},\frac{1}{10},\frac{1}{11},\frac{1}{12},\frac{1}{13},\frac{1}{14},\frac{1}{15}>\frac{1}{16};\ldots$$

Taking into account these inequalities, we get the following for the harmonic sum:

$$\frac{1}{2}+\frac{1}{2}+\left(\frac{1}{3}+\frac{1}{4}\right)+\left(\frac{1}{5}+\frac{1}{6}+\frac{1}{7}+\frac{1}{8}\right)+\left(\frac{1}{9}+\frac{1}{10}+\frac{1}{11}+\frac{1}{12}+\frac{1}{13}+\frac{1}{14}+\frac{1}{15}+\frac{1}{16}\right)+\ldots>$$

$$>\frac{1}{2}+\frac{1}{2}+\left(\frac{1}{4}+\frac{1}{4}\right)+\left(\frac{1}{8}+\frac{1}{8}+\frac{1}{8}+\frac{1}{8}\right)+\left(\frac{1}{16}+\frac{1}{16}+\frac{1}{16}+\frac{1}{16}+\frac{1}{16}+\frac{1}{16}+\frac{1}{16}+\frac{1}{16}\right)+\ldots=$$

$$=\frac{1}{2}+\frac{1}{2}+\frac{2}{4}+\frac{4}{8}+\frac{8}{16}+\ldots=$$

$$=\frac{1}{2}+\frac{1}{2}+\frac{1}{2}+\frac{1}{2}+\ldots$$

And this last sum is clearly divergent.

mathematician in history, with more than 1,500 written articles and more than 500 co-authorships. That record has led to the curious concept of the *Erdős number.* Everyone has a determined Erdős number associated with them; Paul Erdős, as the protagonist of this 'theatrical function', is the only one with an Erdős number 0. Everyone who has carried out a scientific collaboration with Erdős has an Erdős number of 1.

Likewise, anyone who has worked shoulder to shoulder with a scientist having an Erdős number of 1 has an Erdős number of 2. And so on. Lastly, for someone who has no kind of connection with Erdős (that is, the person has not carried out any scientific collaboration with a researcher who has an Erdős number), then we say that the person has no Erdős number, or that their Erdős number is indeterminate (some say that it is infinite).

Much has been said about Erdős numbers in the world of mathematics, and a research project was even created, which can be consulted at http://www.oakland.edu/enp/. When one looks at these numbers, the first thing that stands out is the fact that they are astonishingly low. It has been proved that among all the mathematicians active in the last quarter of the 20th century who have a defined Erdős number, the range of values in no greater than 15 (nearly all of them have an Erdős number lower than 8) and, in fact, the average value of all of them is approximately 5. This value crosses the boundaries of mathematics due to the interdisciplinary nature of science and knowledge. Thus, for example, the linguist Noam Chomksy and the astronomer and writer Carl Sagan have an Erdős number of 4 and, therefore, their non-mathematical collaborators also have a relatively low Erdős number.

What is most extraordinary about this is not so much the fact that the Erdős number of a scientist chosen at random should so often have a low value, but rather that this phenomenon is repeated in numerous contexts which seemingly have nothing in common. Similar phenomena are also to be found in great social networks, in the transmission patterns of some diseases, and in the hierarchy of the Web. Such phenomena are what are known as the *small world effect* as in the colloquial expression "It's a small world!". This notion was studied in-depth for the first time in 1967 by the psychologist Stanley Milgram, though throughout the first half of the 20th century several writers had already begun to speculate on its existence.

The idea behind it is very simple. Suppose we have 100 acquaintances, and that each of them also has a similar number of acquaintances. We say that these unknown persons are at distance 2 from us, while our acquaintances are at distance 1. The number of acquaintances of our acquaintances comes, by using the multiplication principle, approximately (the acquaintances we have in common ought to be taken into account, but here, to simplify, we shall not do so) to $100 \cdot 100 = 10{,}000$ people (each of our acquaintances provides 100 acquaintances of level 2). If we now generalise the argument, the number of people who are at distance n from us should be in the order of 100^n. The growth is thus exponential (and therefore very fast) in the number of individuals depending on the distance they are from us; in this way, with very few steps, a huge range of people can be covered. Readers can try out the number of steps needed to find a link with the President of the United States, with Robert Mugabe or David Beckham... They can be assured that the degree of separation will be far lower than they expected!

To study this issue empirically, Stanley Milgram carried out the following experiment. He asked a number of people on the west coast of the United States to send a packet to another person living in Massachusetts. The difficulty of this request was that those taking part in the experiment did not know the address of the person who the parcel was to be sent to, but they did know their name, their profession, and that they lived in Massachusetts. Milgram told them that the procedure they had to follow was to send the parcel to the person they knew who was, in their opinion, the most likely to know the addressee. That person would have to do the same, successively, until the packet was delivered personally to its addressee. Despite the fact that the participants expected that the task would involve a large number of people, the parcels were delivered to the correct person after, on average, five and a half intermediaries.

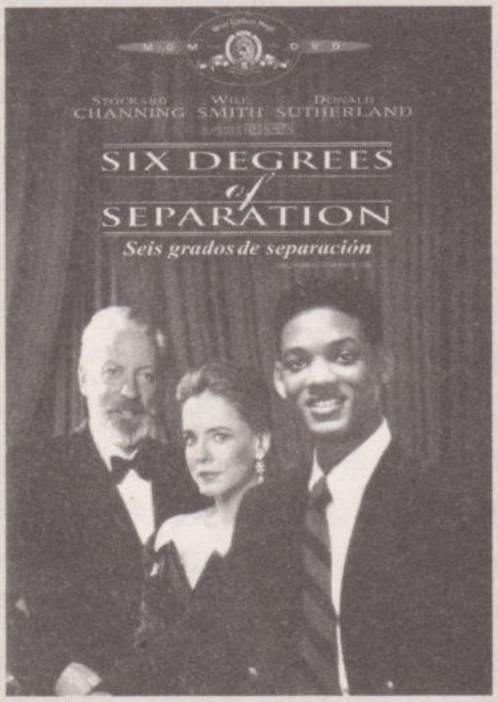
Milgram's discoveries were published in *Psychology Today* and inspired the popular saying of "six degrees of separation" between anyone on the planet. This idea was in fact the inspiration behind a movie called *Six Degrees of Separation* and a TV series named 'Six Degrees', and has caught on in modern-day popular culture together with Andy Warhol's saying "Everyone will have their *fifteen minutes of fame*". A globalised and interconnected world, just as in a graph.

Erdős had his own dictionary to refer to certain aspects; for example, if it were particularly ingenious and beautiful, a proof belonged to 'The Book'. The Book

A PIECE FROM THE FILM *SIX DEGREES OF SEPARATION*

Six Degrees of Separation, a film from the year 1993 directed by Fred Schepisi, cleverly introduces the idea of the small world. In it, Paul (played by Will Smith) manages to infiltrate the world of Ouisa and Flan Kittredge, two New York art merchants, passing himself off as the son of actor Sidney Poitier. When Ouisa discovers that Paul is not who he says he is, she says:

> "I read somewhere that everybody on this planet is separated by only six other people. Six degrees of separation between us and everyone else on this planet. The President of the United States, a gondolier in Venice, just fill in the names. I find it extremely comforting that we're so close. I also find it like Chinese water torture that we're so close because you have to find the right six people to make the right connection. It's not just big names, it's anyone. A native in a rainforest, a Tierra del Fuegan, an Eskimo. I am bound, you are bound, to everyone on this planet by a trail of six people. It's a profound thought. How Paul found us. How to find the man whose son he claims to be, or perhaps is. Although I doubt it. How everyone is a new door opening into other worlds. Six degrees of separation between us and everyone else on this planet. But... to find the right six people."



Poster for the movie
Six Degrees of Separation.

was the place where the gods gathered together all knowledge. Another of his terms was 'epsilon' which he used to refer to small children. In mathematics it is common to use this word in reference to a very small, insignificant amount. And the most important term, 'to die' meant to stop doing maths, while 'to leave' referred to death in the physical sense. Life, for him, was solely an exercise in creating mathematics.

As we shall see in the following chapters, Erdős has not died, he has simply left: his legacy to mathematics is still very active.

Chapter 4

Counting (Without Using Your Fingers)

How dare we speak of the laws of chance?
Is not chance the antithesis of all law?
Bertrand Russell

In the same way that there are jokey stories (perhaps off colour to modern sensibilities) about the activities of people from different countries, the world of mathematics is not exempt from stereotypes and comparisons with other branches of knowledge. One of these jokes, known all over the world in its different versions, goes as follows:

> An engineer, a mathematician and a physicist are staying the night at a hotel. The engineer notices that smoke is coming from the kettle in his room, so he gets out of bed, unplugs it, puts it under the shower and cools it down, then he calmly goes back to bed.
>
> A little bit later, the physicist can smell burning. He gets out of bed and sees that a cigarette end has fallen into a waste-paper bin and some pieces of paper in it have caught fire. He starts to think: "This could be dangerous if the fire were to spread, the high temperatures could kill someone. I ought to put out this fire. How should I go about it? Let's see... I could cause the temperature of the bin to fall below the ignition point of the paper, or perhaps I could deprive the bin of oxygen... I could do that by pouring water over it." So he picks up the bin, goes to the shower, and fills it with water. Then he calmly goes back to bed.
>
> The mathematician realises that his bed is on fire, caused by some embers from his pipe which have set fire to the mattress. This does not surprise him: "It doesn't matter. There is a solution to this problem," and he calmly goes back to bed, as it is gradually engulfed in flames.

The mathematician's attitude to the imminent tragedy reveals something more than just a silly joke. Despite the wit, the story shows a way of thinking and working which goes further than just extinguishing fires.

What we see... and what we don't see

In the world of mathematics it is the norm to be interested in finding certain mathematical constructions (a graph with interesting properties, a set of natural numbers with a special characteristic, and the like) or, alternatively, to prove that an object with these properties cannot exist. In the first case, just showing one specific example is more than enough. But what happens if we are not clever enough to find a way to construct it? Then, unfortunately, we have to make do with showing that an object with those characteristics exists, but without showing an example.

Let's look at a simple example to illustrate this point. Let's suppose that in our son's school classroom a study is carried out and it is found that the pupils' average height is 1 metre 60 centimetres. Just with this knowledge we cannot determine how tall our son is (we cannot even state that he is 1.6 m tall). In spite of that, however, we will be able to make the following categorical statement: "*there is* at least one pupil whose height is greater or equal to one metre sixty centimetres". We do not know who satisfies that property, but we know that someone satisfies it. That fact is clear, because if it were not so, the average value would have to be lower than the value that we have got.

In this chapter we shall see applications of these ideas to discover structures hidden within structures which, seemingly, have no order. We shall discover that 'absolute disorder cannot exist', or rather, that in systems that are big enough there will always be other smaller ones with structure, with some order. These systems will relate to graphs, to points on the plane, and even to numerical sets. And to get a good understanding of these ideas we shall look at pigeon-holes, old school reunions and balloons.

Sharing out breakfast

We shall now look at an elementary principle which can have implications that are more profound than might be expected. Let's suppose that in an attack of generosity we have bought freshly baked cakes for each of our workmates, that is, one per

person. This way, we can be sure that our colleagues will be in a good mood at break time. Unfortunately, our plans are spoilt because we did not take into account that one of them is sick and has not come in to work. There is, therefore, a serious problem, as we have one cake left over. If all the cakes have to be eaten during break time, it will be necessary for someone to make an extra effort and have two.

This is an example of what is known as the *Dirichlet principle*, also known as the *principle of the holes* or *the pigeonhole principle.* The principle is based on the following observation: if N pigeons want to get inside a dovecote with $N-1$ entrances, then two pigeons will have to go in through the same hole. This example with pigeons is essentially the same as our initial example with croissants, but we have swapped birds for baked goods, and entrances to the birds' nest for workmates.

Leaving behind the pigeons, cakes and holes to summarise, the Dirichlet principle tells us the following:

"If we have N balls and we want to place them in M boxes, with $N > M$, there will be one box that will contain at least 2 balls."

Later in this chapter we shall get more complex results than this seemingly elementary principle, but it is important to note that the Dirichlet principle is an existential result. In other words, it assures us of the *existence* of a certain pattern but does not tell us which elements fulfil it. Going back again to the avian version, we know that two pigeons must have slipped into the dovecote through the same hole, but we do not know which ones have done so. That is to say, within a structure we are affirming the *existence* of another smaller structure with a known form. This idea will be key later in helping the reader understand the philosophy of what is known as *Ramsey's theorem* and its implications in geometry, in combinatorics and in number theory.

There are numerous varied applications of Dirichlet's principle. To use the principle it is only necessary to be able to identify who the pigeons are and what the holes are. Let's look at some examples where we can use this methodology. One curious case is the observation that we shall now show. We take an extract from the book *Curso de geometría métrica, tomo* (*Course of metric geometry*) by the great Pere Puig i Adam. This work served as a fundamental reference in most engineering courses in late 20th-century Spain. Overleaf we transcribe the following extract straight from that book:

"There are numerous examples and stories that show the inadequacy and deceptiveness of intuition. For the sake of brevity and simplicity we shall just make do with the two following examples:

1. Supposing we take a person with an average level of education, a person who is aware that Spain has more than 20 million inhabitants and that our scalps have fewer than 5 hairs per square millimetre. And suppose we ask that person if they can be certain there are two Spaniards who have exactly the same number of hairs on their heads. Having no comparative experience in such matters, that person will no doubt immediately reply that there is no way of knowing the answer.

However, a very simple piece of reasoning would enable us get where

DIRICHLET AND THE DAWNING OF THE ANALYTIC NUMBER THEORY

Johann Peter Gustav Lejeune Dirichlet (1805–1859) is remembered for his contributions to number theory. Born into a Belgian family from Richelet (his surname was Lejeune Dirichlet: *le jeune de Richelet,* 'the young one of Richelet'), studied first in Germany and later in France, under the tutorship of one of the greatest mathematicians of the time. After spells at the universities of Breslavia and Berlin, he went on to occupy the professorship at Göttingen that Gauss had left vacant on his death. He married Rebecka Mendelssohn, sister of the famous composer Felix.

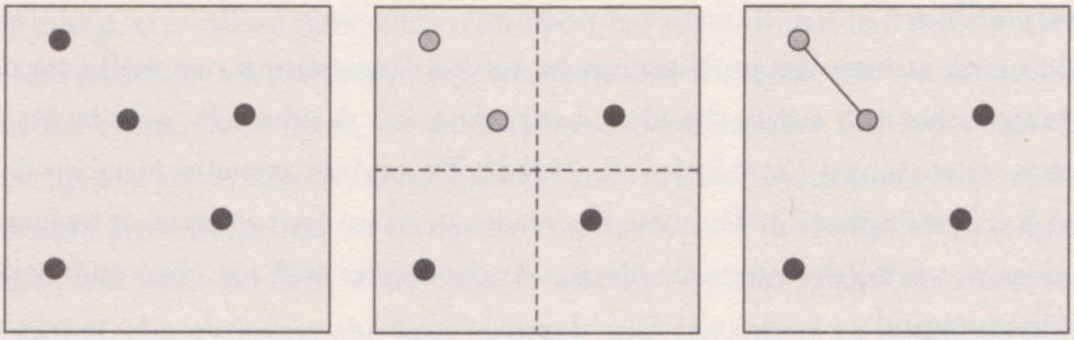
As well as giving his name to the combinatorics principle that we are studying in this chapter, Dirichlet made key contributions to the field of analysis; more specifically, he was the first to use analytical techniques in problems of number theory. That was how, in 1837, he solved one of Gauss's conjectures: Dirichlet's theorem for arithmetical progressions. This result, which we shall look at later, is considered to be the first in analytical number theory.

intuition does not reach, and to answer 'yes', as, if all Spaniards had a different number of hairs on their head, there would have to be one with more than 20 million hairs, requiring a head with a surface area of more than 4 square metres.

2. ..."

Puig i Adam uses the Dirichlet principle indirectly. In this case the pigeons are the Spaniards, while the pigeonholes are the numbers of hairs on the head. The upshot is that a question which at first might seem impossible to answer can be solved by using a very elementary mathematical principle.

Let's look at another application of Dirichlet's principle in a more geometrical context. Suppose we draw five points inside a square of side 1 unit. Note that there will always be two of these points that, at the most, will be at a distance from each other of $\sqrt{2}/2$. To prove it, we can divide the square of side 1 unit into four equal square of sides 1/2 unit.



Five points inside a square of side 1 unit. By the pigeonhole principle, two of them belong to a subsquare of side 1/2 unit. These two points give us the result of the formulation.

We now have the scenario for defining our pigeons and our holes. As we have five points and four squares to put them in, two of the points, in accordance with Dirichlet's principle, must be drawn inside the same square. As the segment of maximum length that can be drawn inside any square is one of its two hypotenuses, those points, which are inside the same square of side 1/2 are separated at most the length of their hypotenuse, whose value is:

$$\sqrt{(1/2)^2+(1/2)^2}=\sqrt{2}/2.$$

Readers who are fans of puzzles and mathematical enigmas can generalise the previous argument and prove that if we draw ten points inside a square with sides 1 unit long, then there are two points whose distance is, at most, $\sqrt{2}/3$. As in the case of the pigeons, we do not know which points are going to satisfy this property, but the principle does guarantee us that such a pair of points exists!

Finally, we shall show an application of the same principle in a more arithmetical context. We shall, in fact, show one of the problems that Erdős most commonly used to spot the mathematical talent of particularly gifted youngsters. One of the many qualities in which Erdős stood out was that of knowing how to assign the right problems to the right people. His interest in detecting potential in maths often caused him to mix with young gifted students. That was how, for instance, during a lunch with the young Lajos Pósa, Erdős came to pose the following question. We have the set $A=\{1, 2, ..., 2n\}$, and let B be any subset of A with $n+1$ elements. Then, in B there are two elements a and b such that a divides b. By the time lunch was over, the young Pósa had found the solution! Years later he frequently worked with Erdős and became a distinguished professor of mathematics in Hungary, his country of birth.

Let's look at how the problem can be solved. To explain it, we shall first show an example that will clarify the ideas. Let's take $n=7$. A subset B could be (as well as many other things) $\{2, 3, 5, 6, 7, 8, 11, 13\}$. This subset contains numbers 2 and 8, and 8 is a multiple of 2. We invite the reader to try to find a subset of 8 elements that does not fulfil the property given... As the reader will see, the result will be very discouraging.

To prove the result, let's write each element of the set A as $2^k m$, where m is an odd natural number and k is greater than or equal to 0 (it will be 0 only if the initial number is odd). As any number is less than or equal to $2n$, the result is that the number m always belongs to the set $\{1, 3, 5, 7, ..., 2n-1\}$. Note that the said set has n elements, and that, therefore, m can be chosen within a total of n elements. As now our set B has $n+1$ elements, by Dirichlet's principle there are two elements of B whose associated value m is equal. The smaller of them will be in the form $2^r m$, and the larger will be $2^s m$, with $r<s$. It is therefore clear that the first one divides the second, just as we wanted to prove.

The Dirichlet principle enables us to solve a wide range of problems concerning the existence of a certain pattern or norm in a superstructure of which we only have partial information. As we shall see in the next section, we can apply the

method in more complex situations, and these ideas will lead to a spectacular result introducing an extremely beautiful branch of mathematics, one whose ramifications span numerous fields – logic, combinatorics, analysis and many others. We are referring to Ramsey's theorem.

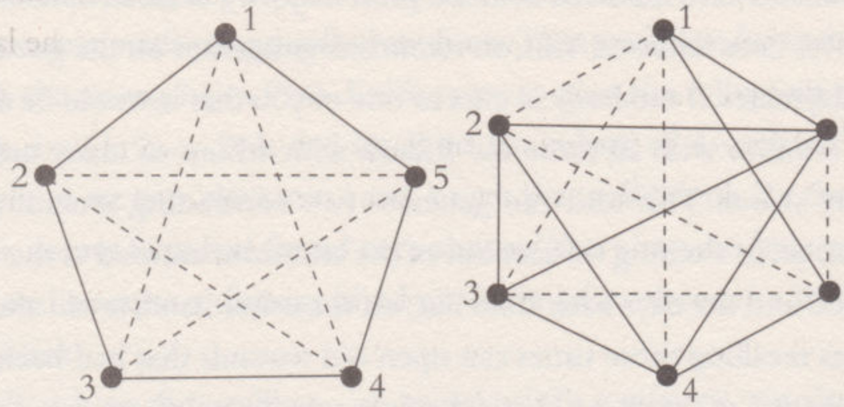
On old school reunions

The reader may find the following situation familiar. One day we receive, out of the blue, news from an old friend from primary school. The pathway through life which our old schoolmates have followed is, in the great majority of cases, different from the one we ourselves have followed. But, on reminiscing together on the good old times, on your old playmates, it suddenly occurs to one of you that it would be a good idea to organise a reunion with all those old friends with whom so many unforgettable moments were shared. This idea can generate two contrasting sensations. On the one hand it can seem exciting on account of the emotions aroused at the thought of one's childhood and the days back then but, on the other hand, it can also be rather unappealing, as recalling those times can open old wounds that had been healed by time and patience. Let's suppose that, whatever our thoughts are, we do decide to attend the event. Once there, we are surprised at the number of people who have turned up, and try to link the freckled little boys nestled in our memories with the overweight great hulks that we bump into on our way to the buffet. What might be thought, and what we shall study next, is whether we can predict anything about the reunion. Has everyone else kept in touch with the others or is nobody able to identify their old playmates. In actual fact, it is not like that.

In the same way that Dirichlet's theorem gives us conditions for existence, in situations like the one described we will also be able to be sure of the existence of certain patterns in totally disordered sets. In the case here, the school reunion, we can state that if more then six people have turned up, then there will *always* be three of them who have kept in touch, or *alternatively* three who have not kept in touch. To explain this, we will again use graph theory language, which enables us to abstract the question and look at just the really important part of the problem. Each person at the meeting will be associated with a labelled vertex (the labels could be natural numbers, for instance). Between any two vertices we shall draw an edge. The graph obtained (each pair of vertices joined by one single edge) is called a *complete graph*. The difference now in respect to what we said about graphs in previous chapters is

that the edges will be labelled or coloured in two different ways: the edge will be solid or dotted depending on whether the persons associated with the corresponding vertices have kept in touch over the years or not. Note that keeping in touch is a symmetrical relationship. If John has kept in touch with Peter, then Peter has kept in touch with John; in other words, the definition of the labelling mentioned above is not ambiguous.

The next figure shows a configuration for a total of five and of six people labelled from 1 to 5, and from 1 to 6, respectively.



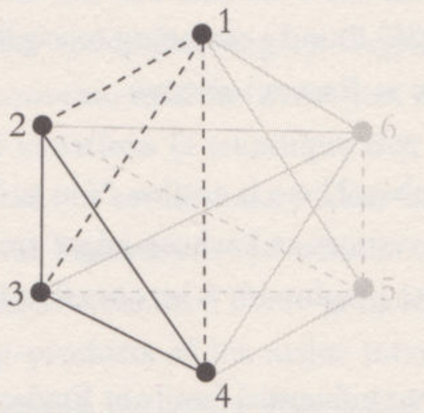
On the left, a distribution of coloured edges in the complete graph with five vertices without monochromatic triangles. On the right, a distribution of coloured edges in the complete graph with six edges. Note that in this second case the triangle with vertices 246 is monochromatic.

After translating the reunion to a labelled graph, the property that we want to prove is this: given a complete graph with more than six vertices and with the edges labelled by two labels (solid and dotted edges), there is always a monochromatic triangle (that is, three edges that define a triangle with the same colour or label). That is because if there are three people who have kept in touch with each other, these three will define three solid edges forming a triangle and, similarly, by dotted edges if these persons have not kept in touch.

Let's go on to prove this. To start with, we shall simplify the problem, by proving, for instance, that the property is true for a group of six people. Note that if we are able to prove the result for a specific case of six vertices, the general result will have been established, as in a complete graph of more than six vertices we can always extract that same structure.

To prove the result, let's suppose that we concentrate on the vertex with the label number 1. That vertex is incident with five neighbours by means of edges (of different colours). The key observation at this point in the argument is that because we have five edges and two different colours or labels, there is a colour that appears at least three times. This statement appears by directly observing all the possibilities or, in a shrewder way, by applying the pigeonhole principle. Let's suppose that the predominant colour is the dotted one, and let's consider the vertices that are joined to vertex 1 by dotted edges. According to the hypothesis there are at least three vertices with this property. Let's suppose, to reduce the argument, that these vertices are those with labels 2, 3 and 4 (and possibly some of the others, but there is no need to consider those); if it were not so, it would suffice to change the labels of the vertices to get that configuration.

Under these hypotheses two things can happen. Either one of the edges between vertices 2, 3 and 4 is dotted, or the three of them are solid. In either of the two cases we reach our goal. In the first one we take the dotted edge; this edge, together with those that join the vertices with vertex 1, forms a triangle of dotted edges. In the second case, as all the edges that are formed with vertices 2, 3 and 4 are solid, in the triangle with vertices 2, 3 and 4 the edges are of the same colour.



The dotted edges are considered here. In this specific situation, the monochromatic triangle is obtained as a result of the triangle with vertices 2, 3 and 4.

We have proved, therefore, that in a complete graph with a number of vertices greater than or equal to six, and with the edges coloured in two different colours, there is always a monochromatic triangle. Note, again, that the argumentation is existential, and that in general we shall not be able to say which triangle will be the monochromatic one.

Let's make the problem a little more complicated. With this result we have shown that if, at our old schoolmates' reunion, more than six people turn up, then there will be three of them who have kept in touch with each other, or who, alternatively, will not have kept in touch.

Can we say something similar for groups of four people (who have kept in touch over the years or, alternatively, have not kept in touch), five people or, in general, for a determined number of people? By translating this problem into the context of graph theory, the question we have is as follows: if we set a value k, is there a value N for which any complete graph with more than N vertices and with edges labelled by two colours contains a monochromatic complete graph of k vertices? This problem, in the case of $k=3$, is one we have already solved: any complete graph with more than six vertices contains a monochromatic triangle. However, is the result true for any value of k ?

As we shall see in the final section of this chapter, we shall be able to say a lot about this result that is completely unexpected, and without constructing any colouring! Despite the fact that elementary methods are not appropriate for such a complex system, we shall see that the basic ideas can still be recycled in a way that is clever enough. The general answer, far from being a mere application of the Dirichlet principle, is the doorway to a new world within the mathematical universe. In 1930 Frank P. Ramsey (1903-1930) found a reasoning to explain this question. This result has been known since than as *Ramsey's theorem*:

"Let k be a whole number. There is a value N in such a way that any complete graph with more than N vertices and whose edges are coloured in two different colours contains a complete graph with k vertices that is monochromatic."

Ramsey was born into an influential family in England in the early 20th century. His father was the president of Magdalene College, Cambridge, and his brother, Michael Ramsey, was to become the Archbishop of Canterbury. As a teenager Ramsey showed an outstanding interest in, and ability for, science, including mathematics. His other fields of interest included economics, psychoanalysis and decision theory. Ramsey went to Cambridge University and from 1924 till his death worked at King's College, where he was appointed under the auspices of the great economist John Maynard Keynes.

Frank P. Ramsey.

Ramsey's scientific work was not extensive due to his early death (aged 26) following complications after abdominal surgery. Despite that, his work was truly groundbreaking. In the economic field he wrote just three articles, in which he covered the calculation of variations (a technique that stems from mathematical physics) and in which he set out revolutionary ideas in theories on taxation and optimal saving. In fact, Paul Samuelson, the winner of 1970's Nobel Prize for economics, says that Ramsey's works are "Three great legacies – legacies that were for the most part mere by-products of his major interest in the foundations of mathematics and knowledge." In his article *Truth and Probability* he developed his own theory of probability, in which he established the principles of the modern theories of subjective probability, utility and decision.

Ramsey kept up a close relationship with the philosopher and father of logical positivism, Ludwig Wittgenstein, and translated his *Tractatus Logico-Philosophicus*. It was Ramsey himself who was instrumental in Wittgenstein getting a position at Cambridge thanks to Wittgenstein's book being accepted as his doctoral thesis.

His contributions to mathematics, and in particular to mathematical logic, were also very significant. In his work *The Foundations of Mathematics* he reconstructed

KEYNES, RAMSEY AND THE KEYNESIAN SCHOOL OF ECONOMICS

Despite his short life, Frank Ramsey's legacy, both in pure mathematics and in other disciplines, is more than extraordinary. It is particularly so in economic theory.

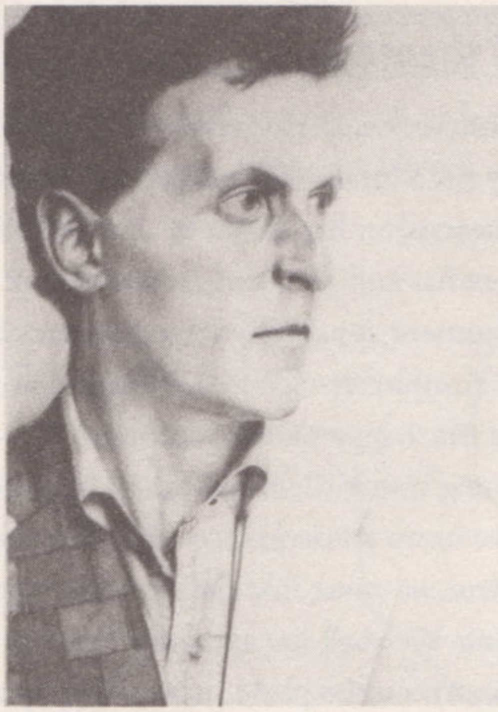
John Maynard Keynes, Baron Keynes of Tilton (1883–1946) is the founder (as the name suggests) of the Keynesian school of economics. This doctrine supports, among other things, interventionist policies by the state (both fiscally and monetarily) with the aim of controlling the problems that appear as a consequence of changes in the economic cycles. He left such a great legacy that he is widely considered to be the father of modern macro-economics. Keynes was aware of Frank Ramsey's intellectual capacity, and said of him:

John Maynard Keynes.

> "From a very early age, about sixteen I think, his precocious mind was intensely interested in economic problems."

In fact, at the age of only 19, the young Ramsey wrote *Mr. Keynes on Probability*, a very critical revision of Keynes' *A Treatise on Probability*. His criticism was so crushing that Keynes had to rework some of his previous arguments.

certain logic theories set out in Russell and Whitehead's *Principia Mathematica*. His interest for the basics led him to try and find the solution to what is known as the *Entscheidungs problem*, an attempt to try and find a mechanical method to decide whether an arbitrary mathematical proposition (belonging to what is called first-order logic) can be proven within a theory or not. In actual fact, in the work *On a Problem of Formal Logic*, he solved one of this problem's particular cases, giving rise to what we know as Ramsey's theorem.



A pencil sketch of Ludwig Wittgenstein, the father of logic positivism and a great scholar of language (source: Christiaan Tonnis).

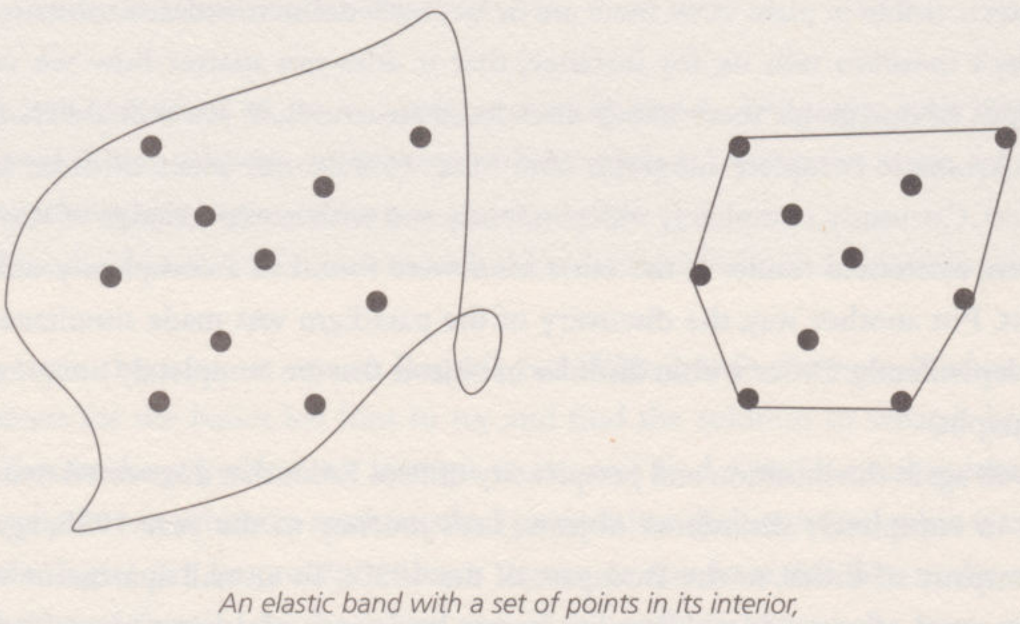
With this result, a new field of mathematics was created, in which order within disorder is studied. This gave rise to a new paradigm, a new 'object of desire' – understanding why, in certain completely disordered mathematical structures, with no pattern visible in plain view, there are in fact well-defined, ordered substructures. Ramsey's theorem tells us, for instance, that it does not matter how we colour the edges of our graph (here is our disordered structure), as we will always find a monochromatic complete subgraph (and here we have our piece of order in the disorder). Curiously, completely independently and with no knowledge of Ramsay's theorem, existential results of the same kind were found in a completely different context. Put another way, the discovery of the paradigm was made simultaneously but independently. Order within disorder in objects that are completely unconnected with graphs.

It was again the intuition and perspicacity of Paul Erdős that discovered structural results in completely disordered objects. Let's journey to the year 1933, again in the company of Erdős, to the Budapest of the 1930s. To a small apartment where a teenage girl, after randomly drawing points on a piece of paper, notes something that will change her professional life, her personal life and her love life...

The 'happy ending' theorem

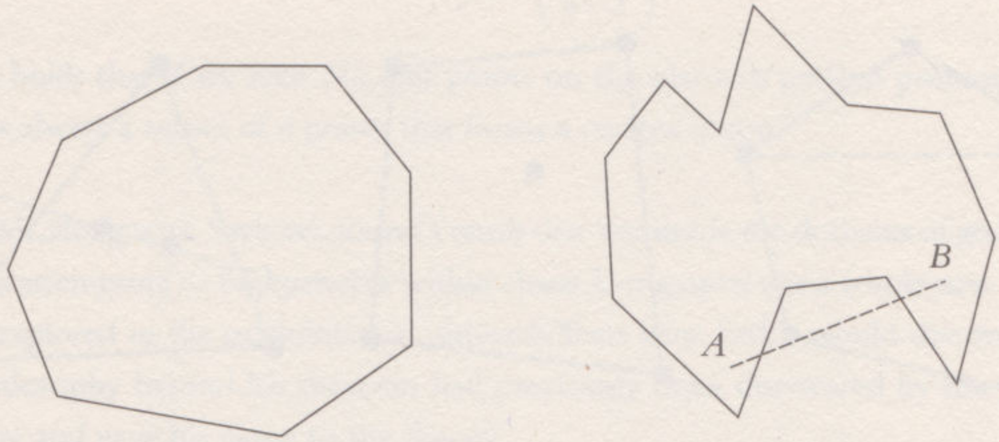
Below the statue of Anonymous, the notary of the glorious King Béla, in that little woodland in Budapest, a crowd of youngsters are having a lively discussion on mathematics. The conversation has arisen by accident, and as a consequence of something a young student has said. At a certain moment in the meeting, she makes an innocent remark, a comment apparently inconsequential but which, though she does not know it, would completely change her future. So much so, that the result would become known as the 'happy ending' theorem.

It was 1933 and the young Esther Klein had almost by accident made an interesting discovery that would have more consequences than she could ever have imagined. To understand this problem we must first define what in geometry is known as a *convex hull* of a set of points. We shall use an elastic band to illustrate this idea. Let's consider a set of points drawn on the plane surrounded by an elastic band. Imagine that the points are anchored to the plane and that we cannot move them. The elastic band is flexible and allows deformations, provided that we do not pass over one of the points. It could be said that the points are the bases of some posts. Let's suppose now that we begin to pull the elastic band tight; there will come a time when the band touches a number of the initial set's vertices. When the band cannot be pulled any tighter, the interior area will be the convex hull of the set of initial points.



An elastic band with a set of points in its interior, before and after being pulled tight.

The convex hull of a set of points has some very interesting properties. For instance, it is always a convex polygon. In other words, it is a polygon with the property that for each pair of the set's points, a line joining them is integrally contained in it. Equivalently, a polygon is convex if each of the angles between its sides is less than 180 degrees.

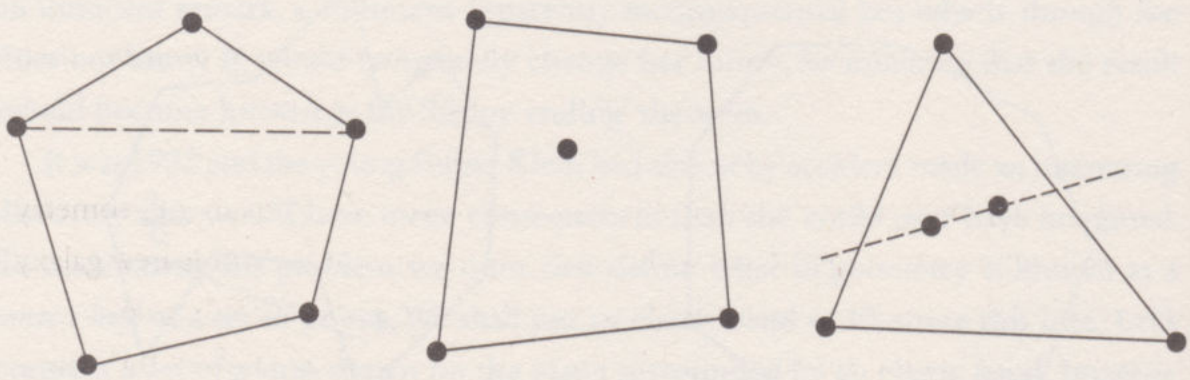


A convex polygon and a non-convex polygon.

The key observation that Esther Klein made was as follows: Take five points on the plane in a general position. This means that no three points are aligned. But apart from that condition, any layout is valid. We shall prove that there is always a subset of four points that define a *convex quadrilateral*. To prove this affirmation we shall use the convex hull of this set of points; the hull may have three, four or five sides. Let's study each of these cases separately:

1. Let's suppose to start with that the convex hull has five points. On account of the pentagon's convexity, to get a convex quadrilateral any subgroup of four points can be taken.
2. Let's now look at the case in which the convex hull has four points. In this case there is no need to present many arguments as the vertices of the quadrilateral are from our initial set, and they already define a convex polygon for us.
3. Finally, let's study the case where the convex hull is a triangle. This is the most complicated situation and will require more work. As a consequence of the hypothesis that we assumed, the two points that do not belong to the elastic band must be in the interior of the triangle. If we consider the straight

line that joins them, then two of the points of the initial triangle are on one side of that straight line (that is so because we assumed the condition that the points are in general position). If we now join the two points on that side of the triangle to the points that are in the interior of the triangle, we get a convex quadrilateral.



The three different situations can occur in Esther Klein's result.

In this way we have proved the result which is generally known as the 'happy ending' theorem:

"In a set of five points on the plane, there are always four of them that form a convex quadrilateral."

As the result mentioned gives no indication of any particularly happy ending, the reader may be puzzled as to theorem's name. We shall therefore continue with the story. Esther Klein's result was easily generalised to pentagons by the fast-thinking, ingenious young Hungarian masters who loved solving mathematical puzzles. But, is it true that given a value n there is another value $f(n)$ for which any set of more than $f(n)$ points in general position contains a convex n-gon? The answer to that question was not at all clear.

Spotting that Esther Klein's observation had opened up new fields in which geometry and combinatorics are combined, Erdős set to work on the question with his friend Georges Szekeres. Erdős foresaw, to a certain extent, that the initial result was simply a very specific case of a richer and much more general theory, which years later would become Ramsey's theorem – finding order within complete disorder.

After several efforts, Erdős and Szekeres managed to solve the problem, and in 1933 came up with the renowned Erdős-Szekeres theorem:

"Let n be a whole number. Then there is a value $f(n)$ with the property

$$2^{n-2}+1 \leq f(n) \leq \binom{2n-4}{n-2}+1,$$

which holds that if we take $N \geq f(n)$ points on the plane in general position, then there is always a subset of n points that forms a convex n-gon."

Erdős, along with Szekeres, found a result that was inside the domains of geometry but was attempting to find patterns within chaos. Lying open was a whole new galaxy to be explored in the mathematical universe. Years later, Erdős would discover that the philosophy behind his theorem had previously been discovered by the young Ramsey, and gave his name to the theory.

A MORE COMPLICATED QUESTION

The natural question is to know whether in the Erdős–Szekeres theorem the value of $f(n)$ is nearer to $2^{n-2}+1$ or to

$$\binom{2n-4}{n-2}+1.$$

There is no known answer to that question. In that same work, in fact, the two colleagues also made the following conjecture:

"Any set of $2^{n-2}+1$ points on the plane in general position contains a subset of n points which form a convex n-gon."

Up to now, the conjecture is known to be true for values of n below or equal to 6, but for other values it is only known that $f(n)$ satisfies the following inequality:

$$2^{n-2}+1 \leq f(n) \leq \binom{2n-5}{n-3}+1.$$

The reader must still be wondering about the name given to Esther Klein's theorem. As a result of the joint work Erdős and Szekeres carried out on the theorem, the latter and Esther Klein ended up marrying, and they shared the rest of their lives together. That is the happy ending referred to by Erdős, who was something of a matchmaker.

George Szekeres and Esther Klein as a married couple, many years after the 'happy ending' theorem came into being. Husband and wife were to die within one hour of each other.

Revisiting probability: the probabilistic method

Up to now we have seen that to calculate the probability of us winning a bet we must use techniques based on combinatorics to work out the number of favourable outcomes. We need to count so as to be able to tell if a determined bet should be made or, on the contrary, whether it is better to hold back. Probability feeds off combinatorics and its techniques when drawing its conclusions. In this section we shall see that in a surprising way we can reach conclusions within the domain of combinatorics – by using probability! Our byword in this section will be "a probable event is a possible event". If we prove that a situation or event has a certain probability of occurring, then the possibility exists. For instance, it is not very likely for it to rain in the Sahara Desert, but the possibility exists, so it may rain.

Let's refine this idea. The most important thing is to remember that all probability is a numerical value between 0 and 1, where probability 0 refers to an impossible

event, while 1 defines the absolute metaphysical certainty. Similarly, if there is an event the probability of which is a certain value p between 0 and 1, then the opposite event will have probability $1-p$. For example, if the probability that it will rain tomorrow in the town we live in is equal to p, the probability that *it will not* rain will be $1-p$, because we can state as with absolute certainty that "either it will rain tomorrow or it will not rain tomorrow".

Bearing that in mind, note that if we prove that a certain event occurs with a probability strictly less than 1, then its complement (that is, that the event will not occur) will likewise have a probability other than 0. More specifically, that complementary event will be *probable* to occur and, therefore, it will be *possible* (as there exists a certain probability, however little it is, that the event will happen). To sum up, proving that something has a probability strictly less than 1 assures us that it is possible for the opposite event to occur (and that, therefore, the possibility exists). This will be the idea that will enable us to be sure of the existence of a determined object with interesting properties without finding an explicit construction of it.

Our objective now will be to combine the enumerative techniques we have looked at previously with this idea, the aim being to obtain certain estimations related to Ramsey's theorem. Let's suppose that we take a complete graph with N vertices and that we colour its edges in two colours, but in the following way: for each edge that we consider we toss a balanced coin (with a probability of 1/2 of getting heads or tails). Depending on the result we get, the edge will be solid or dotted respectively. (That is, according to the result we get from the coin, we draw the edge in one way or the other.) The first question is, using this *random* colouring process, what is the probability of colouring the graph in a predetermined way? Note that the complete graph with N vertices has as many edges as non-ordered pairs of different vertices. This corresponds to subsets of two vertices, and is equal to the combinations of size 2 in a set of size N; put in another way, it is equal to the combinatorial number

$$\binom{N}{2}.$$

Each of these edges is drawn in one of the two colours with a probability of 1/2, in such a way that, in accordance with Laplace's rule, the total probability of getting a predetermined colouring is equal to

$$\frac{1}{2^{\binom{N}{2}}}.$$

In other words: by using this process, all the colourings have the same probability of appearing! (That said, the reader should not be surprised by this, as this phenomenon is the same one that we mentioned for lottery games.)

We shall now do some calculation that is a little more complicated: we shall calculate estimations that a random coloring of the complete graph with N vertices has a monochromatic complete graph with k vertices. Note that we get an estimation by means of the following argument. From the N vertices of the initial graph we choose k of them, which will form the complete graph with k monochromatic vertices. We can make this choice in

$$2\cdot\binom{N}{k}$$

different ways, where the binomial indicates the ways to choose k vertices (not ordered) in a total set of N vertices, and the second comes from the fact that the colouring can be done by using two different colours (solid or dotted). The probability that we choose the same colour for each of the

$$\binom{k}{2}$$

edges that form these vertices will be equal to $2^{-\binom{k}{2}}$. The rest of the edges, which are

$$\binom{N}{2}-\binom{k}{2},$$

can be coloured as we like and, therefore, the result we get when we toss the coin before drawing them does not matter.

For this, the probability that in a random colouring of the complete graph with N vertices we will get a complete graph with k monochromatic vertices is less or equal to the number

RANDOM GRAPHS

There are phenomena in the real world for which a probabilistic analysis is the most efficient method of drawing conclusions. For instance, the dynamics of social networks and the practically instantaneous creation of thousands of web sites – these can only be explained in random terms.

This is the starting point for what are called *random graphs*. There are numerous ways of building a random graph, but one of the most widespread is the model called $G(n,p)$, introduced by Edgar Gilbert in 1957 and based on the following. We take n vertices, and between each pair of them we either draw an edge with probability p, or we do not draw it with probability $1-p$. It may seem simple, but this model has for decades focussed the study of great random problems related to graphs.

There are some very similar models, like the one introduced by Erdős and his Hungarian colleague Alfréd Rényi in 1959. What is most curious about this is that these authors did not realise the implications that their seminal ideas would have in the study of complex phenomena such as computer networks and the spread of epidemics, plus many more! In the words of Erdős himself:

"The evolution of random graphs may be considered as a (rather simplified) model of the evolution of certain real communication-nets, e. g. the railway, road or electric network system of a country or some other unit, or of the growth of structures of inorganic or organic matter, or even of the development of social relations. Of course, if one aims at describing such a real situation, our model of a random graph should be replaced by a more complicated but more realistic model."

$$2\cdot\binom{N}{k}\cdot 2^{-\binom{k}{2}}.$$

Note that we must use the lesser or equal, as we are over-counting (it could happen that in one same colouring we have several complete graphs with k monochromatic vertices). Let's look at this expression in more detail. What property is satisfied provided this value is less than 1? As long as this number is less than 1, we are stating that the opposite event (that is, getting a colouring where there are

no monochromatic complete graphs with k vertices) has a certain probability of happening, and therefore it exists as a possibility (we can toss the coin, colour the edges and get this result). Provided it holds true that the expression above is less than 1, there will exist colourings of the initial graph without monochromatic complete graphs.

It is not very difficult to see that the first value of N for which the formula is greater than 1 (and when we become unable to carry out the thought experiment overleaf) is $N = 2^{k/2}$. This fact proves what is called the lower bound for Ramsey's theorem:

"We need at least $N = 2^{k/2}$ vertices to be sure that in a colouring with two colours of the edges of the complete graph with N vertices there is always a monochromatic complete graph with k vertices."

This result was discovered by Erdős in 1947, thus inaugurating what were to be the probabilistic methods of combinatorics. In fact, the critical value for which Ramsey's theorem starts to hold true is called the *Ramsey number*, commonly denoted by $R(k)$. (We have already seen in this chapter that $R(3) = 5$.)

The first method that we have been through here is the first step towards a series of techniques in which the aim is to get constructions, both in combinatorics and in geometry, or in number theory, in which the fundamental element is probability. That's why these existential constructions are based on what it known as the *probabilistic method*. It is so important, both in theory and in applications in diverse branches of mathematics, that there are two top-level magazines specialising in these ideas: *Combinatorics, Probability and Computing* and *Random Structures and Algorithms.*

One result in which this probabilistic method is used, without being an explicit construction, also appears in the context of graph theory. Its technical details are more complicated, and we will make do with explaining the result. First let's look at some definitions. A graph's *chromatic number* is the minimum number of colours needed to colour the graph's vertices in such a way that those that are joined by an edge with a different colour. The *girth* of a graph is the length of its shortest cycle. For instance, the complete graph with n vertices has chromatic number n (as all the vertices are joined to all the others and, therefore, we need n colours to draw the points), while its girth is 3 – any three vertices define a triangle, and this is a cycle of minimum length.

AND...WHAT ABOUT THE EXACT CALCULATION OF THE RAMSEY NUMBER?

Together with the bound obtained by Erdős it is known that

$$2^{k/2} \leq R(k) \leq \binom{2k}{k}.$$

Calculating the exact value $R(k)$ is a problem that is not at all easy due to the huge number of calculations that have to be made. Here, for instance, the bound given above gives us, for $k=5$:

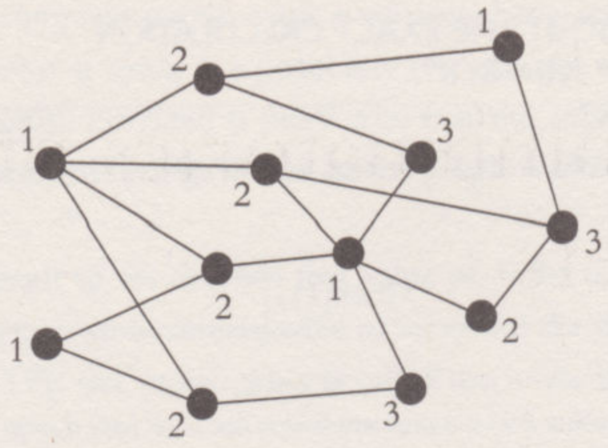
$$4 \leq 2^{5/2} \leq R(5) \leq \binom{10}{5} = 252.$$

Let's suppose that we want to prove whether a determined integer n that is in this range is equal to the Ramsey number $R(5)$. We would have to prove that any colouring of the edges of the complete graph with n vertices has a monochromatic complete graph with 5 vertices, and that there is at least one colouring of the edges of the complete graph with $n-1$ vertices without a monochromatic complete graph with 5 vertices. The number of calculations that have to be made for the graph with n vertices will be in the order of

$$2^{\binom{n}{2}},$$

as that is the number of possible colourings. That number is, for values of n that are not actually all that big, astronomical (for $n=100$ already comes to approximately $2^{10000/2}$), so the problem will not admit a direct algorithmic solution (the number of operations required is completely unmanageable for a computer). Erdős himself had sensed the difficulty of the problem. He once expressed his doubts about whether a solution would ever be found by making the following remark:

> "Suppose aliens invade the Earth and threaten to obliterate it in a year's time unless human beings can find the Ramsey number $R(5)$. We could marshal the world's best minds and fastest computers, and within a year we could probably calculate the value of $R(5)$. If the aliens demanded the Ramsey number for $R(6)$, however, we would have no choice but to launch a pre-emptive attack."



A graph with chromatic number 3 (the labels show the colour) and girth 4 (it has no triangles, but it does have a cycle of length 4).

Note that if a graph has a lot of edges, then many of the vertices will be connected to one another and, therefore, the graph's chromatic number will be very high. On the other hand, as it has a lot of edges it will be very simple to find very short cycles that will give us a small value for the girth (just as happened in the complete graph). Reciprocally, if the graph has very few edges, the chromatic number should be very small, as few of the vertices are joined. And, just as we have reasoned already, the graph's girth will be high. There is, then, a dichotomy involving the girth of a graph and its chromatic number: high chromatic numbers mean small girths, and small chromatic numbers give rise to large girths. The natural question is whether we can construct graphs with both large girths and large chromatic numbers.

This result was proven by Erdős in 1959 using more advanced probabilistic methods, but with the same seminal ideas: to prove that an object exists it is enough to prove that it can occur with a positive probability. Thus, the Erdős chromatic number theorem states that:

"For any two values *r* and *s*, there is a graph whose chromatic number is greater than *r* and whose girth is greater than *s*."

The difficulty in explicitly constructing a graph that fulfils those two properties (a high chromatic number and a high girth) is still an ongoing unsolved problem. As we can see, there are many occasions in which we have to be satisfied with just knowing a solution exists to be able to sleep easy – even though the house may still be on fire.

Chapter 5

The Combinatorics of Numbers

Order is a home's most beautiful ornament.
Pythagoras

There is another type of combinatorics that concerns whole numbers and their properties, a combinatorics that could also be described as elementary since the problems that are formulated in it stem straight from the most basic definitions. This branch of mathematics, which is halfway between number theory and combinatorics, is usually called *number theory combinatorics* or the *additive number theory*, and it was again the multitalented Paul Erdős who laid the foundations for this discipline. It is an area in which there is a great deal of activity in science, largely thanks to the results and enigmas that we shall explore in this chapter.

We shall start by referring to the prime numbers, which are without doubt the building blocks on which arithmetic is based. We shall continue with representation functions, which will be, to a certain extent, the basic combinatorics objects that will codify much of the combinatorics information for us. We shall relate this concept to old and still unsolved enigmas, such as the well-known Erdős–Turan conjecture. Later we shall make further use of Ramsey's theorem to approach other combinatorics problems within the discipline of additive number theory. In particular, we shall see that back in the year 1927, Van der Waerden obtained a result in Ramsey theory, but in the context of figures. It is surprising that the first results in Ramsey theory were produced independently by diverse authors and in diverse areas, but all in a common historical framework. *Van der Waerden's theorem* is just the visible tip of a much bigger and deeper iceberg. We shall explain what the density of a numeric set is. We shall also see that starting from certain very general conditions we can find out a lot about the internal structure of that set. We shall look at what is known as *Szeméredi's theorem*, the greatest result in combinatorics in the second half of the 20th century, and we shall finish off by showing the *Green–Tao theorem*, proven in 2003 by Ben Green and Terence

Tao, for which the latter was awarded the Fields Medal at the 2006 International Congress of Mathematicians in Madrid. But before we arrive at the Olympiad, the Fields Medal and the amazing results obtained in modern mathematics, we should start off in the catacombs, in other words, by taking another look at those who were the building blocks of arithmetic.

The fundamental pieces of arithmetic

Without fear of error, we can confidently state that the most desired and most studied objects in mathematics are the prime numbers. They are simple, natural and very easy to define, but inside them they hold surprising mysteries of the mathemati-

THE MOVIE *CUBE* AND 'ASTRONOMICAL' CALCULATIONS

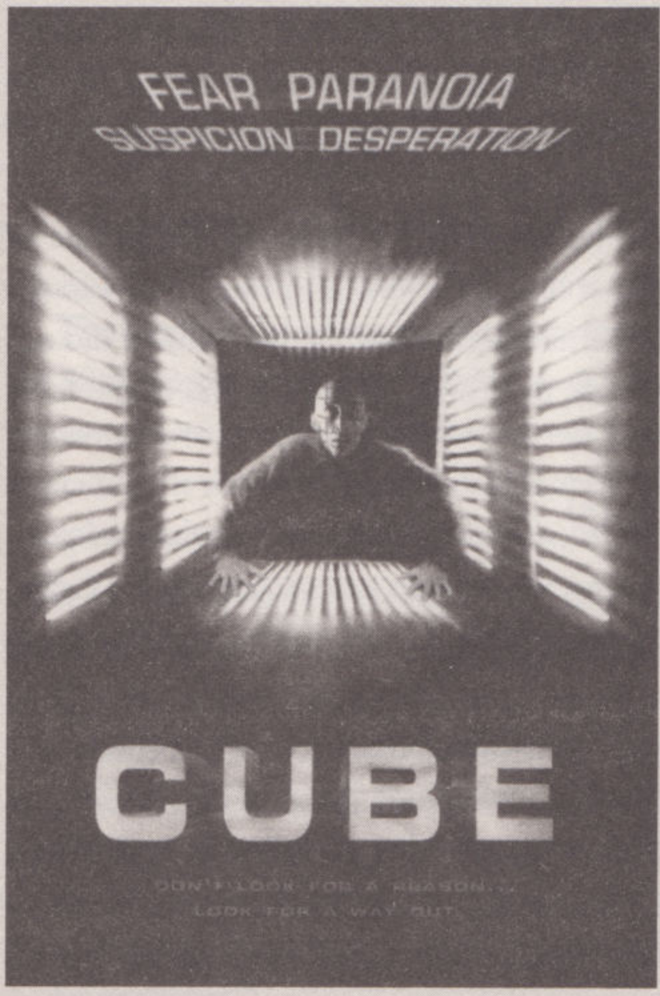
In the claustrophobic film *Cube* (directed by Vincenzo Natali in 1997), a group of strangers wake up in a futuristic cubic room. To survive they must find a way out by moving to adjoining rooms, which turn out to be identical to the first one. As time goes by they discover that some rooms contain lethal traps. Leaven, a maths student, works out the pattern to determine which rooms are full of hazards. On every door he discovers three numbers of three digits. If any of them is a prime number, then the room contains a danger that they must avoid at all costs. So, finding out if a room is safe or not depends on searching for prime number in each of the numbers that they find on the door to the room. Unfortunately, Leaven points out that it is going to be very hard for the survivors to make a decision:

> "... Look! No one in the whole world could do it mentally! Look at the number: 567, 898, 545. There's no way I can factor that. I can't even start on 567... It's astronomical!"

Note that every three-digit number is less than 1,000 and that $\sqrt{1{,}000} = 31.62\ldots$ so to find out if a number containing three digits is prime or not it is just necessary to know if it is divisible by number 2, 3, 5, 7, 11, 13, 17, 19, 23, 29 or 31. Mentally, the task is difficult and requires practice, but in no case is it astronomical! Particularly in our case, where it can easily be seen that 567 is divisible by 3; 898 is divisible by 2 and 545 is divisible by 5.

cal paradigm. According to the German Leopold Kronecker, "God created natural numbers; all else is the work of man." The prime numbers therefore encapsulate the essence of mathematics.

As the reader will recall, a natural number is prime if its only divisors are 1 and itself. It is a simple and natural condition. Primes are the bricks with which arithmetic is built and, therefore, it is essential that they be understood. The first issue to be dealt with is knowing how to detect a prime number. In other words, given an integer n, do we have enough criteria to decide if it is a prime or not? At a theoretical level, the question is trivial as we only have to prove what the divisors of n are, and that can be done simply by carrying out all the possible divisions and checking that the remainders are always other than zero. However, for practical purposes, deciding



Poster of Cube. *The 1998 Sitges Fantasy Film Festival awarded this film two prizes for best film and best script.*

if a determined number is prime or not is a complicated computational problem. Imagine if we had to calculate the divisors of a number of, for example, a million digits. Even modern IT equipment cannot manage computations like that! In fact, problems of this type tend to be complex due to the fact that computer capacity for calculations is limited when the arithmetic involves gigantic integers. These very limitations are put to good use in encryption systems that allow transactions to be carried out safely over the Internet.

The next issue in this study deals with the general formulae that will find all the prime numbers for us, or at least some of them. In the same way that the expressions n^2 and $7n$ generate results (by replacing the variable n with any natural number), where all the numbers are squares and multiples of seven, respectively, is there a mathematical formula that for each entry of n produces a prime number for us? The answer is no, and that is due to the fact that prime numbers behave in a very mysterious way. It could even be said that they behave randomly.

Since finding a general formula for the prime numbers is not feasible, the next natural question is to ask how many primes exist below a given number n. As before, there is no general formula for this problem. But, curiously, what are known are approximate results, and these achievements have been precisely the speciality of many of the greatest mathematicians in history.

The story goes that in 1792, when he was only 15 years old, the 'prince of mathematics', Carl Friedrich Gauss (1777–1855) made some surprising observations: despite the fact that the quantity of prime numbers below a given number did not appear to follow any mathematical formula, their order of growth was of the form

$$\frac{n}{\ln(n)}$$

By order of growth we understand that the difference between the real value and the one that the said formula predicts is much less than the value n (it could be said that it is a very good approximation). Thus the said function is a very good approximation of the real function. That a 15-year-old should reach this conclusion just with numerical calculations done by hand is truly amazing! In fact, on account of how difficult it was, he stored this problem away in his attic of unsolved problems. Years later, when the question was by then in the public domain in the world of mathematics, Gauss again surprised the scientific community with the story from his teenage years. He had foreseen it many years before.

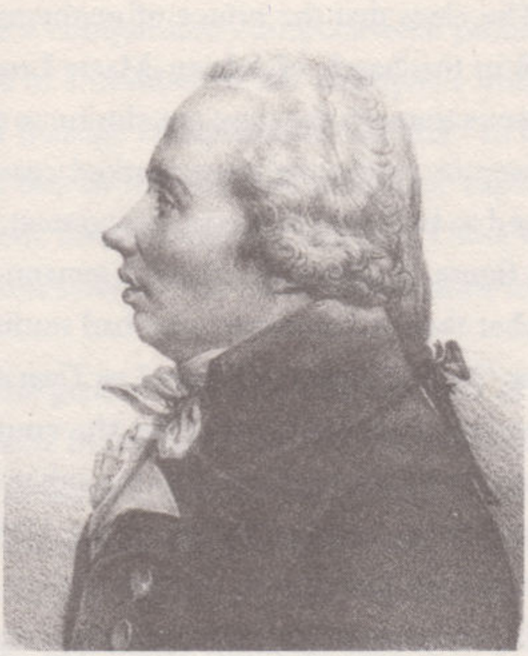
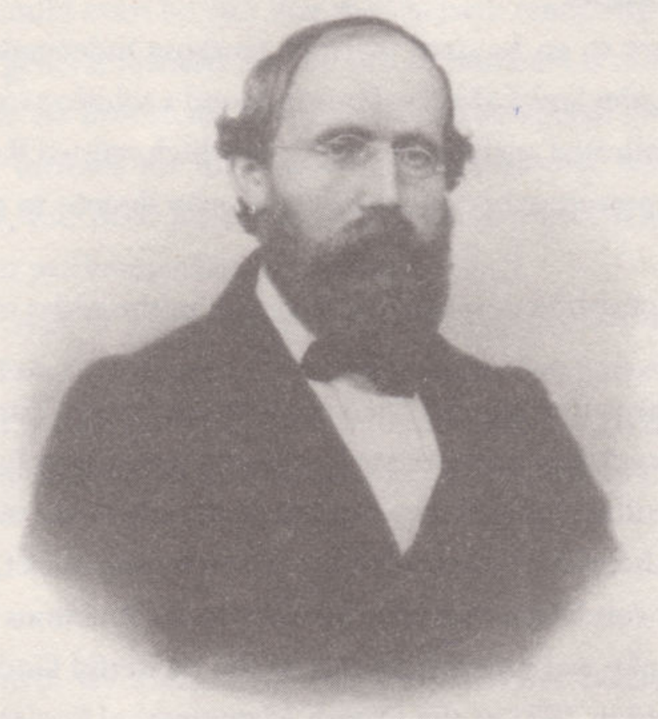
The clues that the prince of mathematics had glimpsed became a conjecture in 1798 in the hands of Adrien-Marie Legendre (1752-1833), who used numerical evidence to reach the same conclusion as the young Gauss had seven years previously. Numerous mathematicians worked ceaselessly on this problem, but the solution refused to be found. It was in this context, research into the number of primes below a set figure, in which Bernhard Riemann (1826–1866) carried out his only work in number theory. Riemann, who had studied under Gauss, in 1859 wrote his famous paper *On the Number of Primes Less Than a Given Magnitude* in an attempt to provide an analytical technique to solve the conjecture. Riemann did not manage to find an answer to the problem. On the other hand, his work would be one of the most important on number theory in the 19th century, since it introduced possibly the most famous function in number theory – the Riemann zeta function. That led to the formulation of the most important mathematical conjecture of the 20th century, the Riemann hypothesis.

Some years were to go by until in 1898, working independently, both Jacques Hadamard and Charles Jean La Vallée-Poussin found a solution to the problem. Both of them used complicated analytical techniques which refined Riemann's profound work. Legendre's conjecture became the *prime number theorem*, in the following form:

"The number of prime numbers less than n is in the order of $\frac{n}{\ln(n)}$."

The proof of the prime number theorem that was known during the first half of the 20th century used very sophisticated analytical techniques linked to Riemann's renowned zeta function. The difficulty was linked directly to the technique's analytical limitations. It was therefore believed that the theorem was, in fact, equivalent to very complex analytical conditions that were inherent to the functions they were working with. That philosophy was superbly set out by the masterful Godfrey Harold Hardy at a conference held in 1921 at the Danish Mathematical Society in Copenhagen.

> "No elementary proof of the prime number theorem is known, and one may ask whether it is reasonable to expect one... If anyone produces an elementary proof of the prime number theorem, he will show that these views are wrong, that the subject does not hang together in the way we have supposed, and that it is time for the books to be cast aside and for the theory to be rewritten."

From top to bottom and left to right: Gauss, Legendre and Riemann. For an entire century, these figures led the search for the distribution of prime numbers.

In 1948 the world of mathematics was struck dumb when Paul Erdős announced that he and another great mathematician, Atle Selberg (1917-2007), had found a truly elementary proof of the prime number theorem in which only basic properties of

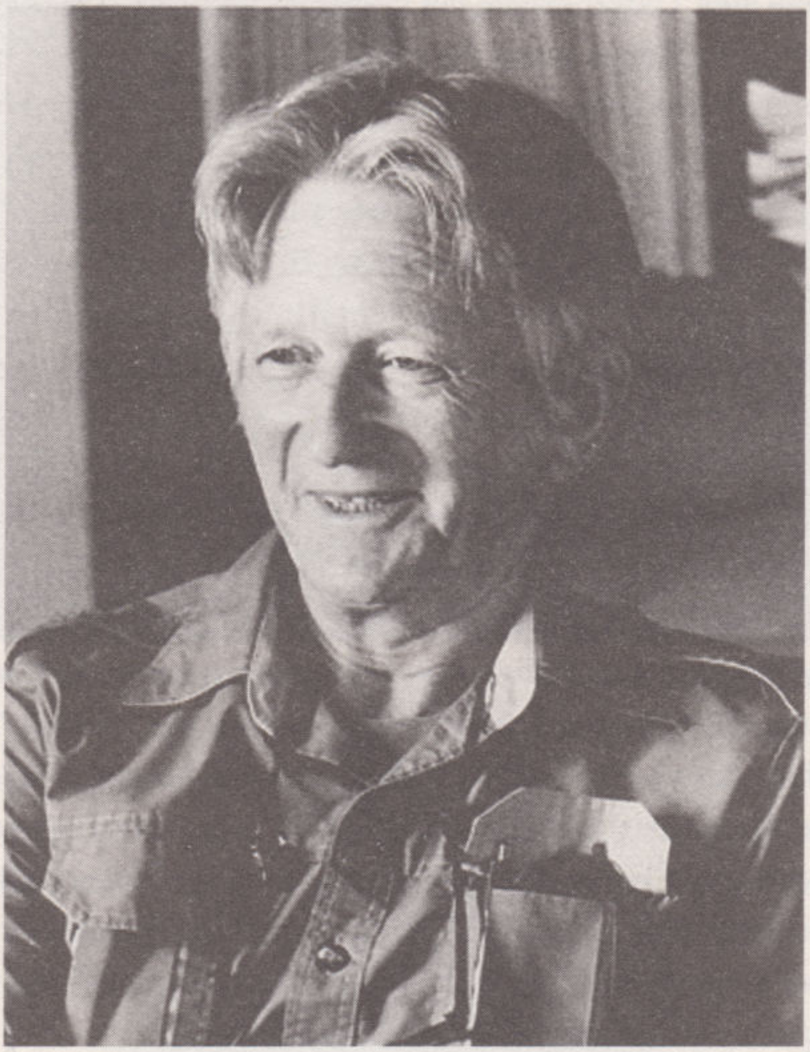
logarithms were used. Unfortunately, this important discovery was eclipsed by events beyond mathematics. If there is one thing that a scientist wants more than to solve a seemingly unsolvable enigma, it is to know that his or her name will live on forever beside that immortal theorem. And so the prime number theorem caused more than just a couple of headaches for our friend Paul Erdős, and it led him to be the protagonist of one of the most heated controversies of 20th-century number theory.

A really awkward situation came about over a result that Selberg had obtained but had not published. In that theorem, the Scandinavian genius had obtained an alternative elementary proof of Dirichlet's theorem (known as the arithmetic progression theorem, which we shall come back to later). Despite the fact that his results had not been made public, Selberg explained the details of the proof to the Hungarian mathematician Paul Turan, who at that time was visiting Princeton.

Following that discussion, Selberg then left Princeton bound for Canada for work reasons. During the time that Selberg was away from the Institute, Turan took it upon himself to offer a public seminar at the Institute of Advanced Study explaining the new ideas that Selberg had told him. Fate and coincidence caused Erdős to be in the hall that day and, consequently, to witness his friend Turan's talk. The shrewd Hungarian mathematician realised that with this new and elementary approach the prime number theorem could be tackled in a completely novel way, and with this in mind he quickly set to work.

Selberg returned from Canada to find Erdős completely bound up in solving the enigma. It was Erdős who remarked to Selberg that, according to his sharp intuition, there existed a still undiscovered pathway leading to the paradise of the prime number theorem. Selberg, who was aware of the possibility, answered Erdős that his intuition was too optimistic. But nothing could have been further from the truth: Selberg also sensed that he was on the right track to finding the elementary proof of the prime number theorem. He, like Erdős, had the intuition to see beyond the problem and spot elegant and elementary solutions to difficult problems. Selberg was a lone researcher, with hardly any collaborators, while Erdős was the champion at gathering scientific colleagues around him.

Not many days went by before Erdős found the way to refine Selberg's result and, in addition, to find the method to prove the prime number theorem through elementary techniques. At last, Erdős had managed it, after more than 200 years of research into the matter.



Atle Selberg, author of the 'fundamental lemma' which led to the elementary proof of the prime number theorem.

Render unto Caesar the things which are Caesar's: Erdős contacted Selberg again to give him the good news. These events were not to the Norwegian mathematician's liking; quite the opposite, in fact. Erdős had already made his findings public and the theorem was already considered to be his. That did not please the shy mathematician at all, so when Erdős suggested to him that they should publish a joint article, Selberg flatly refused. They had completely different outlooks on work, which was reflected in the fact that each of them published their proofs of the theorem (Selberg managed to adapt his method to solve the enigma completely independently of that used by Erdős). It was his proof of the prime number theorem (as well as his crucial contributions to the analytical number theory) for which Selberg was awarded the

Fields Medal in 1950. Erdős was never rewarded with that distinction, although he did later receive a distinguished prize in the mathematical world – the Frank Nelson Cole Prize from the American Mathematical Society in 1951, an important award in number theory.

Representation functions

We shall now look at a type of combinatorics problem that is very different from those we have seen up to now. The problem that we shall begin to discuss is seemingly simpler than those we have looked at in the previous chapters, but an expert combinatorialist is used to meeting problems that are easy to formulate but impossible to solve.

We shall start with a very elementary problem, one which could be explained to children and which they would no doubt be able to solve! Given the set of natural numbers $A = \{0, 1, 3, 4, 5, 8\}$, in how many ways can we write a determined natural number as the sum of the two elements (possibly the same one) of the set A? For example, 0 can be written only as $0+0$; 1 can be written as $1+0$, or $0+1$ (note that it does not matter what the *order* of the addends is); 2 can be written only as $1+1$; and 8 can be obtained from $3+5, 5+3, 8+0, 0+8$, or as $4+4$. By using more sophisticated mathematical language, we say that the number of ways to represent a natural n as the sum of two elements of the set A is the *number of representations* of that natural number with respect to A, or also the *representation function* associated with A. In our example, it holds that the number of ways of representing 0 and 2 is equal to 1 ($0+0$ and $1+1$, respectively), the number of ways of representing 5 is 4 ($5+0, 0+5, 1+4$ and $4+1$), while the number 15 does not admit any representation and, therefore, its representation function is 0.

The problem we have posed is simple, as the set A is finite. Since A contains a number of elements that we can count on our fingers, it therefore has a maximum element, and any value greater than double that maximum cannot be represented as the sum of two elements. In our example, the number of representations of 16 is equal to 1 ($8+8$), but for values of n greater than 16 the representation function will be 0, as the elements of A are too small to be represented. So then, the interesting problem comes from studying the representation function when A is a finite set. And precisely this consideration, i.e. that there exists an infinite number of elements, in our case turns a trivial problem into a devilishly difficult one.

We shall start off with a drastic simplification of the problem. Is there any set of whole numbers A, i.e. an infinite set, made up in such a way that any number is written as the sum of only two elements. Clearly there isn't: if it were so, the element 0 would have to belong to the set so as to represent 0. In the same way, 1 must be an element of the set so as to represent 1, and then $1+0$ and $0+1$ are two representations other than 1. Therefore we have proved that a set A with this property cannot exist. In the same way, similar reasoning proves that there cannot exist an infinite set A in which the number of representations is the same for all natural numbers (that is, having any other number instead of 1).

This suggests to us that we should try to set less stringent conditions in the formulation so as to be able to get a better understanding of the nature of representation functions of infinite sets. Let's consider the following problem: is there an infinite set A that has a representation function that is constant from a determined point onwards? In this case we allow some liberty so that in the small values the number of representations take any value, but that from one point to infinity the number of representations is constant.

This problem now ceases to be trivial and needs a bit of ingenuity. The first solution to this problem dates from 1941 and was down to Erdős and one of his most fervent collaborators, Paul Turan. Surprisingly, his reasoning makes use of a very sophisticated technique, what is known as *Fabry's theorem* for defining functions using a circle. The curious thing about it is that an exceedingly simple proof exists that confirms that this situation cannot occur.

The key to the problem consists of observing that for an element a of the set A, the number of representations of $2a$ is an odd number, while the number of representations of $2a+1$ is an even number. Given that there are infinite elements in the set, we can conclude that the representation function takes even values and odd values an infinite number of times and, therefore, it cannot be constant from one point onwards.

And how can we go about proving this fact? By using combinatorics, of course! Let's first look at the number of representations of the natural number $2a$. Note that a is an element of the set, and $a+a$ is a valid representation. The crucial observation is this: if there is any other representation of $2a$, the addends from which it is composed must necessarily be different. So, for example, if $b+c=2a$, with b different from c, then the representation $c+b$ is different from the representation $b+c$, such that the number of ways of representing the double of a must be an odd number – namely

the only representation in which the two addends are equal, plus an even number of representations in which the addends are different. On the other hand, we do not get a term where the addends are repeated when we consider the representations of $2a+1$, as this number is odd. Therefore, any representation of this number has to be carried out by using different numbers and, in short, the number of representations will be an odd number.

SIDON SETS, OR, AS ERDŐS DESCRIBED THEM, THE ADDITIVE NUMBER THEORY

Probability feeds off combinatorics, combinatorics feeds of graph theory, and knowledge is interconnected by totally unexpected pathways. And the same things happens with combinatorial number theory. In this case, one of the theory's key problems stems from a branch that is seemingly so distant from combinatorics – harmonic analysis.

The story goes that in 1932 the Hungarian analyst Simon Sidon asked the young Erdős about the arithmetic of the frequencies of certain periodic functions. What is interesting about the problem is that, irrespective of its origin, the background to the puzzle came from an issue that mixed arithmetic and combinatorics. The property that the eccentric professor suggested to the young student fascinated the genius because of the problem's combinatorial flavour, which made it one of Paul Erdős's favourites – to such an extent that he often went back to studying it in his works.

A sequence of integers is said to be a *Sidon set* if any pair of elements in the set, whatever the order, produces different sums. For instance, the set $\{1, 4, 7, 10\}$ is not a Sidon set as $4+7=1+10$. There are infinite questions concerning these sets, such as obtaining good estimations for the number of elements of a Sidon set with all the elements smaller than a given integer. Curiously, research into Sidon sets can be used in technical areas such as the design of radars and communications systems – unexpected applications for multidisciplinary theories.

In view of the result that we have just proved, the reader might wonder how to generalise the context of this question. If, instead of asking for the number of representations to be always the same, we set the condition that it can vary, but within a pre-determined margin (let's say, for example, that the number of representations can only be 1, 2 or 5), then the problem ceases to be an exercise of ingenuity and

becomes one the most important open problems in combinatorial number theory. It is again one of Uncle Paul and Turan's conjectures, also from 1941, and whose solution carried a prize of $500; the *Erdős-Turan conjecture:*

"Let A be an infinite sequence of natural numbers. If the representation function is other than 0 from one point onwards then the representation function cannot be bounded."

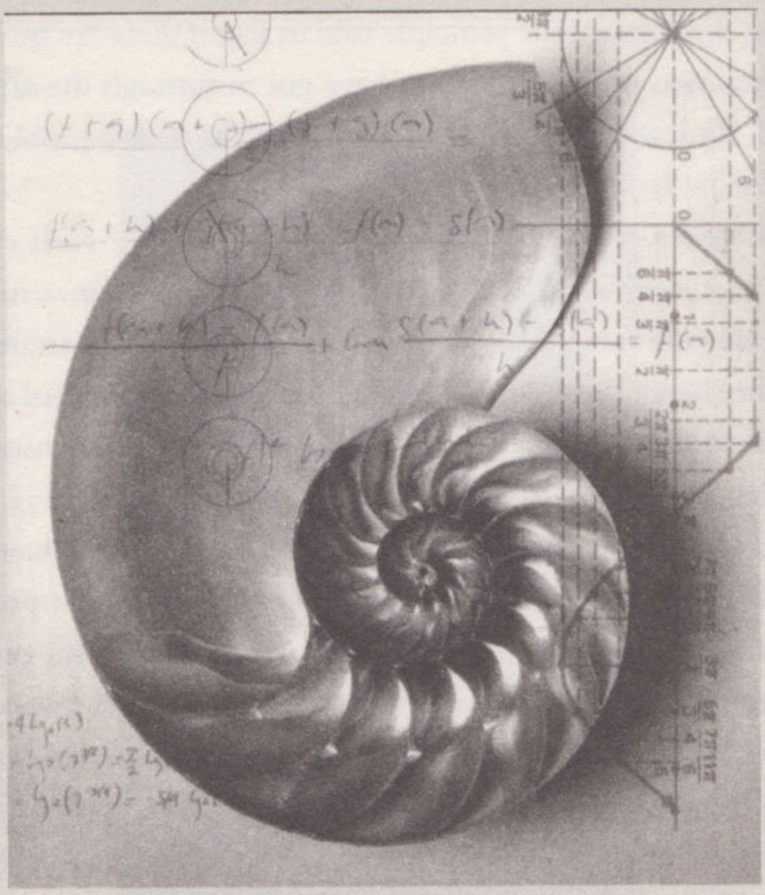
Let's try and understand what this conjecture is telling us: Let's take an infinite set A in such a way that any sufficiently large natural number can be written at least as the sum of two elements of A. Put another way, any sufficiently large natural

A CONJECTURE WITH A LONG HISTORY

In 1742, mathematician Christian Goldbach wrote to the great Basle scientist Leonhard Euler with the aim of throwing light on a pattern that he had observed. It seemed that all even numbers could be written as the sum of two primes. Euler was not able to solve that enigma, and as years went by this humble observation became a key issue in number theory.
Goldbach's conjecture is simply one more problem in the world of representation functions. Little progress has been made towards a complete proof of this result. Powerful computers and advanced computational methods have managed to verify that the conjecture is true for very large numbers but, up to now, no argument has been produced to allow the problem to be completely solved.
There are some very interesting questions related to Goldbach's conjecture to which the answers are known, however, among them Vinogradov's theorem and Chen's theorem. The former proves that there is a point forwards of which any large odd number can be written as the sum of three prime numbers. In the latter theroem, Chen managed to improve the result and prove that any sufficiently large even number can be written as the sum of two primes, or as the sum of a prime plus a semi-prime (that is, the product of two odd prime numbers). In this second result it is not possible to specify when each case occurs.
Unfortunately, the analytical methods used to prove these two results do not imply a satisfactory proof of Goldbach's puzzle, which is nowadays the subject of intense research carried out by the greatest modern-day mathematicians.

number admits a representation as a sum of two elements of A. Then the conjecture tells us that under these hypotheses, the representation function cannot always keep below 100, or 1,000, or 10,000: there will always be a value (it may be immense) for which the number of representations will be greater than the predetermined value.

Beyond the innocence of this problem there exists a question with opposing interests. In the same way that we saw that a graph with a high chromatic number must give rise to a small girth and vice versa, in the numerical context in which we find ourselves a similar, but more complicated, problem appears. The probabilistic method allowed for a margin to be found in which the two values (chromatic number and girth) could be as high as we wanted. In our situation, the dichotomy is as follows: if the set A has many elements, then it will be able to represent most of



Reproduction of the cover of an edition of the novel, by the Greek writer Apostolos Doxiadis, Uncle Petros and Goldbach's Conjecture, *in which a retired mathematician suggests to his nephew that he solve the famous problem.*

the natural numbers, even at the cost of having a lot of representations of each large number. Reciprocally, if we want to have few representations (or better, a bounded number of representations) we will have to choose sets A without too many elements, with the risk that there may be large values that cannot be represented.

Little progress has been made since Erdős and Turan formulated this conjecture in 1941. The only thing more that is known for the moment is that if a sequence A fulfils the hypotheses of the Erdős–Turan conjecture, then the representation function must have some value above 7. It's still a long way before we arrive at the infinite!

Weight is important

Let's continue studying the infinite sequences of natural numbers, but now from a somewhat different point of view and by returning to the Ramsey theory. Remember that the pigeonhole principle told us that if there are more pigeons than pigeonholes, then two of the birds must have got in through the same hole. In this section we shall look at a similar question, but with an added problem: the number of pigeons will now be infinite!

Suppose we take a set of natural numbers and we draw each element (that is, each number) in a colour. Of course, our palette will contain a finite number of colours, so we are colouring an infinite set of elements using a finite set of colours. We might be very lucky and find that the proportion of numbers of one determined colour is similar to that of another colour, or quite the opposite might occur. Now, however, the first thing that the reader must note is that there will be at least one predominant colour (though it is true that there may be more than one). In other words, there has to be one colour whose proportion is large compared to the total. If it was not so, all the colours would appear few times, and this cannot happen if we are going to colour all the whole numbers! In line with this observation, the fact is that on colouring the natural numbers with a finite number of colours one of them will be very similar to the initial set (that is, it has a proportion that is high with respect to the initial set). To refine this idea we shall use some very disciplined objects which will give us an idea of the type of structure that we are studying – arithmetic progressions.

Let's review the concept. An arithmetic progression of length r and of difference b is a sequence of numbers in the form $a + bk$, where k takes values

between 0 and $r-1$. Those objects, the arithmetic progressions, are to arithmetic what monochromatic complete graphs are to graph theory – the underlying substructures of the completely disordered objects that we are studying. In natural numbers there exists a very large number of arithmetic progressions, and they are of all possible lengths. So if we want to specify that when colouring with a finite number of colours there is always a colour whose structure is very much like that of the initial set, what we will have to do is to show some property that is transferred from one set to the other.

This topic is involved in the work of the Dutch algebraist Bartel Leendert van der Waerden (1903-1996), which is a key element in Ramsey's theorem applied to set combinatorics. According to Van der Waerden's theorem,

"If we colour the natural numbers with a finite number of colours, there will be a monochromatic set with arbitrarily long arithmetic progressions."

A photograph of a young Van der Waerden, the discoverer of the theorem that bears his name.

Note that this theorem states that there is a colour (we do not know which beforehand, as the theorem only guarantees its existence) that will be very structured. It will contain arithmetic progressions that are as long as we want (of length 5, 50 and so on). This theorem is, in fact, a very specific case of a much more general result involving natural numbers and arithmetic progressions. The key condition in Van der Waerden's theorem is that in the colouring we use a *finite* set of colours and, in fact, this condition is too severe in many cases. A consequence of this is that there is at least one colour whose proportion in respect to the total whole numbers is high. It is what we shall call the *density* of the set, and is what we shall deal with now.

The key is that in sequences of numbers with a lot of elements we shall be able to find some very special hidden substructures. With the concepts given in Chapter 4, in which we managed to hunt down hidden substructures in completely unordered objects, we shall define the notion that will enable us to specify the idea of proportion within an infinite set. To do this we will have to return to probability and remember the Laplace rule. But here there is an additional difficulty – we will have to deal with an infinite number of favourable cases and possible cases.

Let's start with the simplest case. Suppose we want to see the proportion of a finite set of natural numbers in relation to a whole set of positive whole numbers. In this situation, it is obvious that the initial set will have a negligible weight, as the total set is infinite. (Here, and analogously to the probabilistic simile, the favourable outcomes are the number of elements of this finite set, whereas the number of possible outcomes is infinite.) It could be said that the weight of a finite set is 0 in respect to the total, as its size in relation to the complete set of numbers is negligible. (A finite number of objects is nothing compared to an infinite number of them!)

Let's make the problem a little more complicated by taking a set with infinite elements, and more specifically by taking an infinite arithmetic progression; for instance, the arithmetic progression 0, 5, 10, 15, 20 and so on into the infinite. If, from among all the infinite ones possible, we randomly pick a number, what probability will we have that the number belongs to that set? That arithmetic progression is infinite and, therefore, we cannot apply the classic method of Laplace's rule, which is the only way we know how to calculate probabilities.

Our intuition, however, tells us the following: any number is a multiple of five, or, alternatively, when we try to divide it by 5 it generates a remainder that is equal

to 1, 2, 3 or 4 (or 0 if it is a multiple of 5). Therefore, what is left over after dividing by 5 gives a way of classifying the integers. So, on choosing a random number, once out of every five times we will get a number that is a multiple of 5. Intuition tells us, then, that a fifth of the natural numbers are of this type and that, therefore, the density of this set is 1/5 (in other words, a fifth of the natural numbers are multiples of 5).

How shall we formalise this idea? Laplace's rule is not applicable in our case here due to the infinite favourable and possible cases, but what we can do is to take all the elements from 1 up to N (with N being very large) and perform the count. So, let N be a large whole number. Now the number of possible cases is in fact $N/5$, so when we divide it by N we get the required density. This value is not dependent on the N that we take (the quotient is always 1/5), so in this case the density is well defined. The same argument works for proving that the density of an arithmetic progression of a difference equal to m is $1/m$.

This is the philosophy that must be used when we want to calculate the density of a more complicated infinite set. Since we cannot apply Laplace's rule, as the favourable cases and the possible cases are infinite (and we do not have much practice in the arithmetic of the infinite), the best way to work generically with weights is by truncating our sets (or, put another way, by choosing a value for N and carrying out the calculation solely for the integers below that N). The density will be, precisely, the value that we get when we make the value of N infinitely large.

We now have the key idea for understanding the formalisation of the following result, Szemerédi's theorem, named after its discoverer:

"Let A be a sequence of natural numbers with positive density. There are, then, arbitrarily long arithmetic progressions formed by elements of A."

This marvellous result tells us the following: if our infinite numerical set is heavy enough in respect to the total (technically, if the set considered has a positive density), then we will always find m elements in our set A (where m can be as big as our imaginations allow it to be), which will be forming an arithmetic progression. Once again, the philosophy of hidden structures within chaotic objects vigorously makes its presence felt.

The path followed to arrive at this result was one of the most important achievements in mathematics in the second half of the 20th century, and it shows once again that mathematics is one single body, with no closed off, stagnant compartments.

The first step towards resolving the problem was taken by the British mathematician Klaus Roth in 1956 when he managed to solve the problem for the case of arithmetic progressions of length 3 (that is, if A is a sequence of natural numbers with positive density, then it contains infinite arithmetic progressions of length 3).

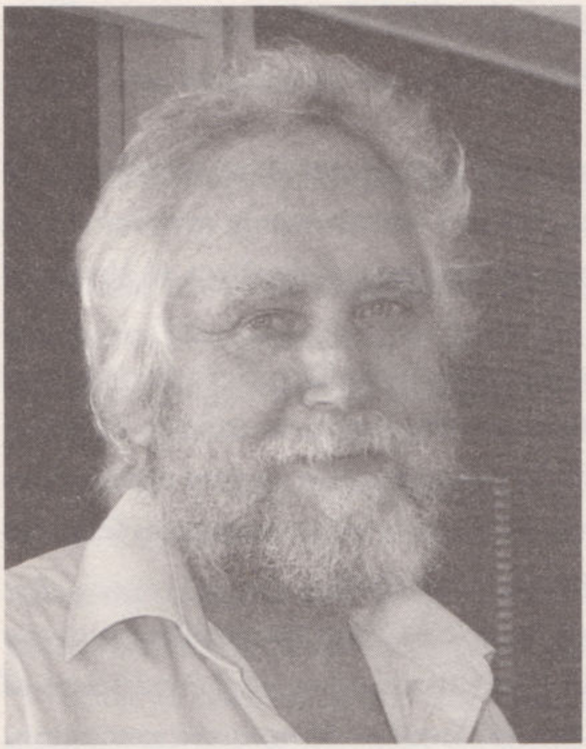
Years later, by using purely combinatorial methods, the Hungarian Endre Szemerédi managed to prove the result for arithmetic progressions of length 4. He would go on to solve the general problem in 1975 by using advanced combinatorial techniques together with a great deal of ingenuity. After that proof, in 1977 the Israeli mathematician Hillel Furstenberg found a new argument to explain the problem, but this time using completely different techniques based on what are known as *dynamical systems*. These are classed within the field of what is termed *ergodic theory*. Finally, in 2001 the British mathematician Timothy Gowers found a third proof of the result by using analytical techniques. The surprising thing about this story is that in many cases proofs of one and the same result produced by different techniques do not mean that the underlying problems are different. Expressed in another way, it is possible for the same ideas to be expressed in a different language depending on the researcher's specialised field. In Szemerédi's theorem, just the opposite occurs: The three proofs use completely different ideas and explain the same phenomenon in completely different ways. In other words, there is no direct way to translate the conclusions of Szemerédi's combinatorial results to Gowers' analytical arguments.

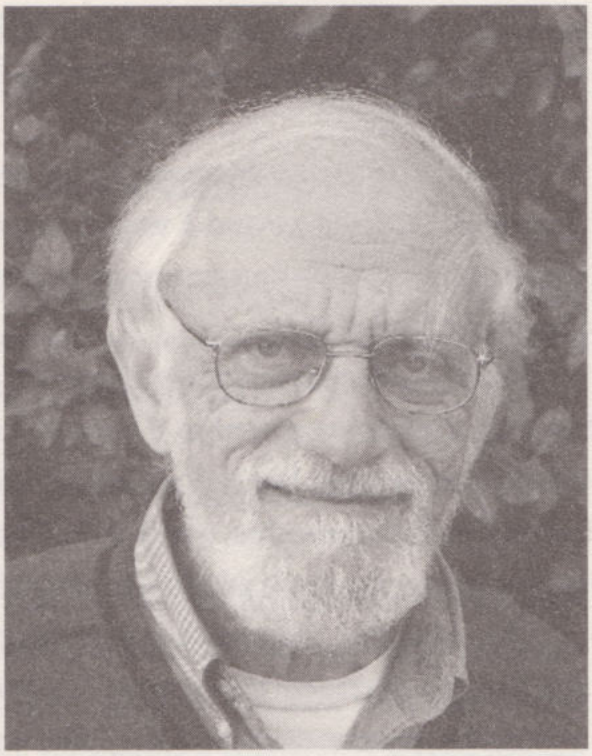
As we shall explain later, far from being a problem, this variety of arguments was precisely what Ben Green and Terence Tao made use of when they proved their theorem, the now renowned Green-Tao theorem on arithmetic progressions, a theorem that again involves the most important personalities in the mathematical kingdom – the primes.

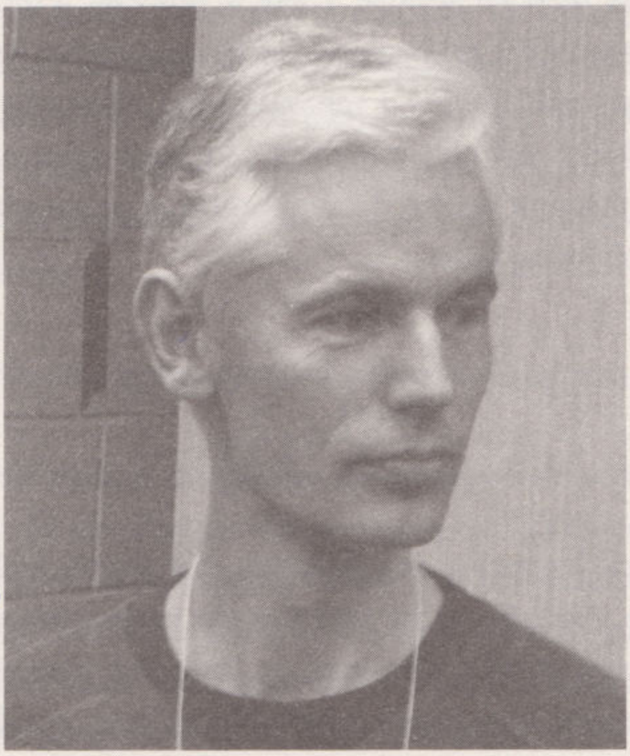
...Despite everything, there are arithmetic progressions

The prime numbers have always been a source of very interesting problems, either for being the basic pillars on which arithmetic is supported, or by being objects with very enigmatic properties. There are problems related to the prime numbers in a great many areas of mathematics. In this last section we shall relate prime numbers to combinatorics.

Let's return to the dispute between Erdős and Selberg. As we mentioned before, the result that was the origin of the race pitting one colossus against the other was what is known as Dirichlet's theorem on arithmetic progressions. This result had







Endre Szemerédi (top left) and Hillel Furstenberg (top right), who together with Timothy Gowers provided three proofs for one theorem.

been conjectured by Gauss, but until 1837 Dirichlet did not find the explanation to it. The formalisation of *Dirichlet's theorem on arithmetic progressions* is very simple:

"Let a and b be whole numbers with no non-trivial divisor in common. Then there are infinite prime numbers in the form $a+bk$."

Note that for the theorem to be true, a and b must not have any divisor in common; otherwise, a and b would be multiples of a certain whole number m and, therefore, by extracting a common factor, all the elements of the form $a+bk$ would also be multiples of m.

There is a deeper and more complicated issue than what Dirichlet's theorem tells us. This explains that within a certain arithmetic progression an infinity of prime numbers may exist, but it does not tell us if they are very near or very distant from one other. Put in another way, it would be useful to know what structure is followed by the prime numbers *within* a given arithmetic progression.

One idea that the reader might think of for approaching the problem is to use the Van der Waerden theorem: it is a result that can assure us of the existence of arithmetic progressions and it applies to general families of numeric successions in which the density must be positive. There is, however, a serious problem when calculating the density of the prime numbers. In our case we do not have the exact formula, but we can apply the good approximation offered by the prime number theorem. Expressed in a different way, for a very large value of n (which is when the prime number theorem works) the proportion of prime numbers compared to the total is of the following order:

$$\frac{\frac{n}{\ln(n)}}{n}=\frac{1}{\ln(n)}.$$

Since the natural logarithm is a function that grows with n, it follows that when we move the n to infinity, we find that the proportion of primes compared with natural numbers shrinks! The need to limit the proportion is 0 and for n large enough, the inverse logarithm is negligible. In short, the density of the prime numbers is equal to 0.

We have, then, a very serious problem, as we are not under the hypotheses of Szemerédi's theorem and therefore we cannot state that there exist arbitrarily long arithmetic progressions in the prime numbers. But, despite everything... can we expect that the primes to contain arbitrarily long arithmetic progressions? There

is numeric evidence that this is so. In 2008 the first arithmetic progression of 25 primes was found, with the general formula

$$6{,}171{,}054{,}912{,}832{,}631 + 366{,}384 \cdot 223{,}092{,}870 \cdot k,$$

where the value of k varies between 0 and 24 (and the resulting values are all prime numbers). Two years later, in 2010, the result was improved up to 26 primes in arithmetic progression by means of the following formula:

$$43{,}142{,}746{,}595{,}714{,}191 + 23{,}681{,}770 \cdot 223{,}092{,}870 \cdot k,$$

where the value of k now varies between 0 and 25. All these results show that in order to find arithmetic progressions of primes we have to start with very large numbers.

What is curious about this issue is that, despite the fact that the prime numbers do not satisfy the conditions of Szemerédi's theorem, its conclusions are still satisfied; put another way, the hypotheses of Szemerédi's theorem are not satisfied for prime numbers, but the conclusion is still true! The problem has recently been solved by two great mathematicians: Ben Green and Terence Tao. What is indisputable is that these two researchers will mark out many of the routes to be followed in mathematics research in the 21st century. According to the *Green–Tao theorem*:

"There exist arithmetic progressions of arbitrarily long prime numbers."

In the words of the authors themselves, the strategy followed when reaching the result was based on the fact that the three existing proofs of Szemerédi's theorem are completely different (we mentioned before that one cannot be obtained from the other by reduction, but rather that they are completely disjointed arguments). It was for that reason that when Ben Green and Terence Tao came across an impasse in their arguments when using a certain technique, they carried on by simply changing to using the tools of another of their works! In this way, by combining the three worlds (seemingly very distant from one another), Briton Ben Green (born 1977) and Australian Terence Tao (1975) managed to solve this great enigma in 2003, without doubt one of the most important in modern mathematics.

Some final words should be dedicated to giving an outline of the biography of

one of the discoverers of this result, Terence Tao. It could be said that this researcher satisfies all the requirements needed to personify the image of a precocious genius in mathematics. Tao is a renowned professor at the University of California, Los Angeles (UCLA), who has worked on a wide range of topics, from differential equations to number theory, and spanning combinatorics and ergodic theory, while touching on more applied matters such as signal theory. The reader may be tempted to think that such a prolific author must be nearing the end of his career. Nothing could be further from the truth. At a very young age Terence Tao was already showing his capacity for mathematics. In fact, he started secondary school at the age of seven, and at nine he was already attending university.

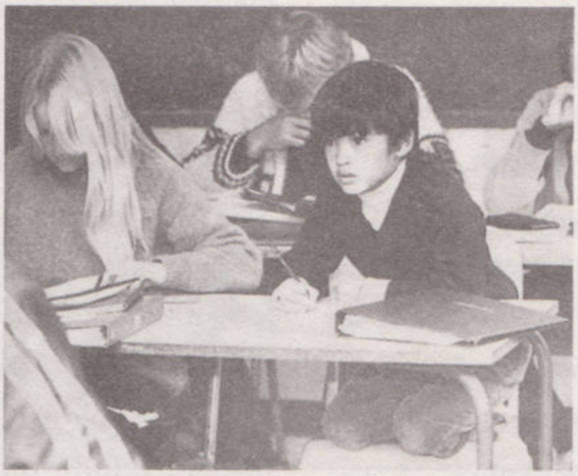
His visit to the International Mathematical Olympiad was astounding. In 1986, at the age of ten, he became the youngest ever contestant. But on top of that he won the bronze medal – all at an age when most children are just beginning to learn the most elementary principles of arithmetic. In the two following years he won a silver medal and then the highest distinction in the competition, the gold. The olympiad is for secondary or high school pupils who are not enrolled at university and his feats were therefore even more astonishing bearing in mind that the students competing had an average age of 17 or 18.

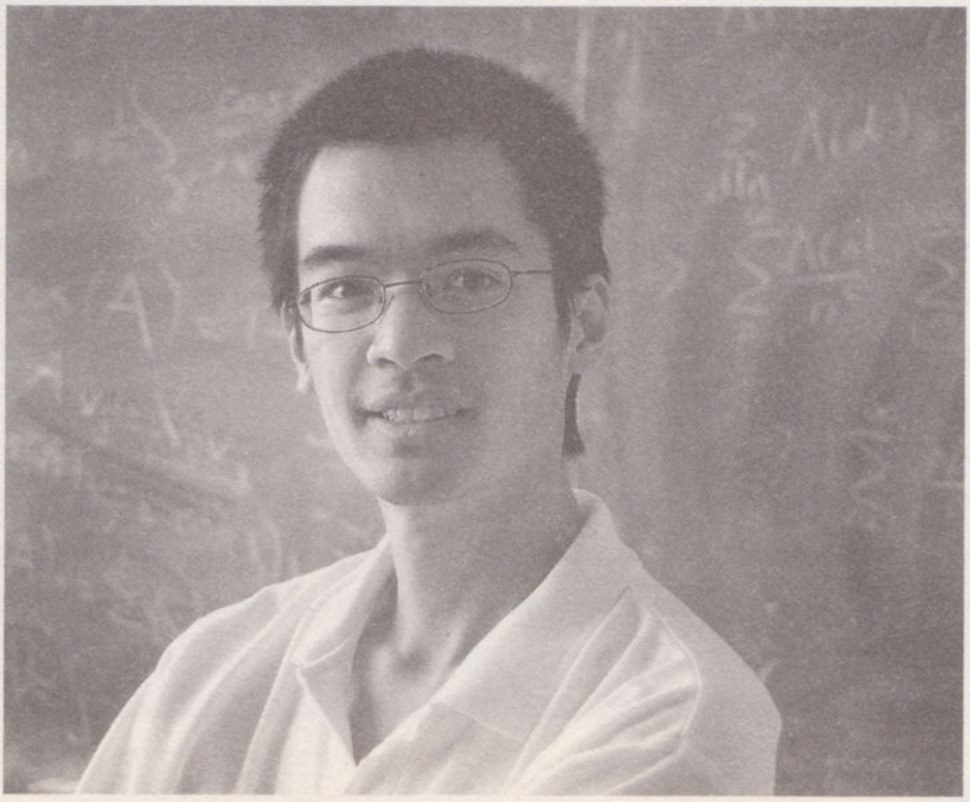
Following these achievements, Terence started at the university in Adelaide (the city of his birth) and was awarded a master's degree at 17 years old. After that he emigrated to the United States in order to study for his doctorate. He received his doctoral thesis at 21 at Princeton University under the direction of Elias Stein, and at 24 became the youngest professor at the University of California, Los Angeles. A few years later, after receiving numerous distinctions for his contributions to science, he was awarded the Fields Medal at the 2006 International Congress of Mathematicians held in Madrid, the greatest aspiration of all young mathematicians.

His capacity for resolving difficult problems is already legendary. In fact, Charles Fefferman (born 1949), who was another child prodigy in mathematics and also a Fields Medal winner, says of him:

> "Such is Tao's reputation that mathematicians now compete to interest him in their problems, and he is becoming a kind of Mr Fix-it for frustrated researchers. If you're stuck on a problem, then one way out is to attract the interest of Terence Tao."

Terence Tao and his colleagues (including Ben Green) are laying the foundations for new mathematical theories in which combinatorics, number theory and analysis are equally mixed. In years to come we will see what results can be proven with all this powerful machinery.

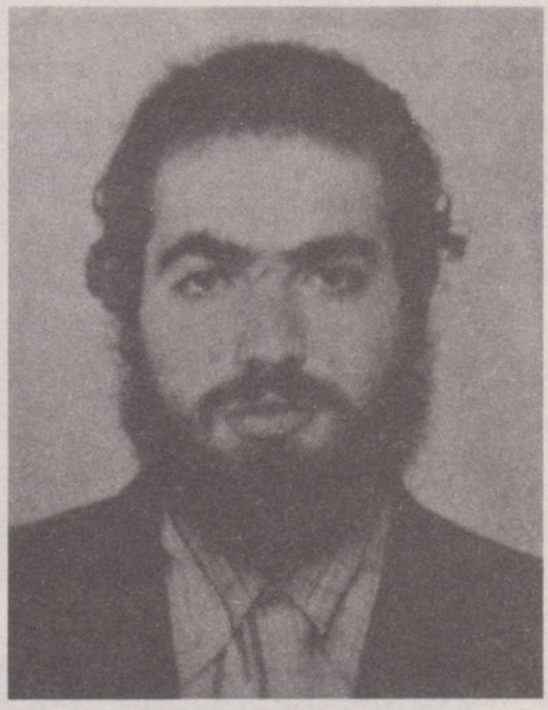




Terence Tao at the age of seven and in adulthood.

THE MOST NOTABLE ABSENTEES AT THE INTERNATIONAL CONGRESS OF MATHEMATICIANS

Shouts of jubilation were heard in the hall when one of the speakers announced "The Poincaré conjecture has been solved!". There was good reason for the excitement, bearing in mind that the problem had withstood more than 100 years of attempts to conquer it. Poincaré's conjecture, the outstanding problem in mathematics in the 20th century, had at last been solved, and the hundreds of attendees at the opening session of the 2006 International Congress of Mathematicians in Madrid had been witnesses to it.



Grigori Perelman.

The proof was based on an arduous and exacting program designed earlier by the American mathematician Richard Hamilton. His strategy, however, was not free from difficulties, and even Hamilton was not able to completely solve them. Some years were to pass before a Russian mathematician, Grigori Perelman, managed to clear up these difficulties and reveal the solution to an enigma that had baffled the whole mathematical community for a century. That feat earned the Russian genius the Fields Medal, the highest distinction in mathematics. But Perelman did not accept it. Beside Terence Tao, Andrei Okounkov and Wendelin Werner, the enigmatic Russian's chair remained empty. Perelman neither came to receive the prize nor accepted the cash reward that was available to him. He refused it, he said, as a protest over the research community's behaviour. There were many who regarded him as a madman, others, a revolutionary, but in any case his contribution to mathematics was incommensurable.

The end

In this book we have aimed to show the way around different areas in combinatorics such as they are understood today, and under the baton of a very special conductor. As the reader will have noted, this discipline is not an isolated area of mathemat-

ics, nor a conglomerate of stagnant problems. There is nothing further from reality. Combinatorics stands out as a discipline in its own right from which the majority of disciplines in the art of exact sciences feed. In all books on combinatorics worthy of the name, there has to be a section on 'open problems' so that those readers who are sufficiently motivated can at least devote some of their spare time to trying to find the answer to unsolved mathematical enigmas.

Let's return to Erdős's conjecture, which we introduced in Chapter 3. Recall that Erdős conjectured the following result, known as the *Erdős–Turan conjecture*:

"Let $A = \{a_1, a_2, a_3, \ldots\}$ an infinite set of natural numbers. If the sum

$$\sum_{i>0} \frac{1}{a_i}$$

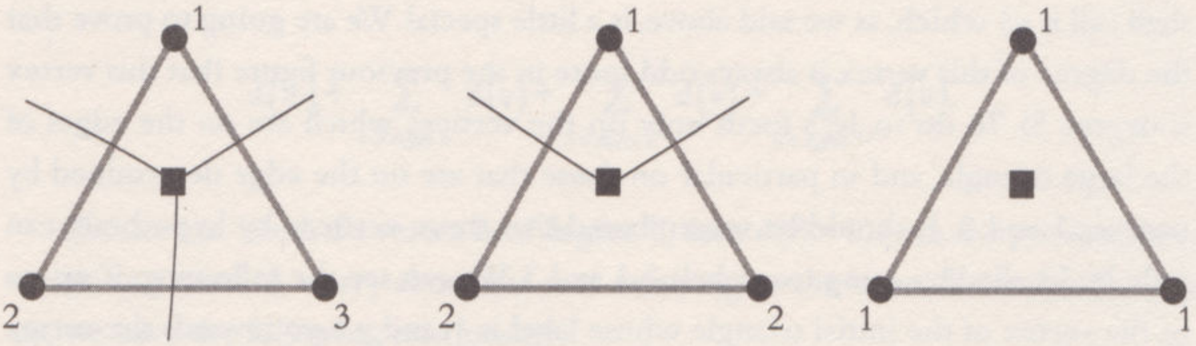
is divergent (that is, infinite), then A contains arbitrarily long arithmetic progressions."

It was for this conjecture that Erdős offered a large sum of money, a total of \$3,000. This seems to indicate that Erdős believed the problem to be extremely difficult, so much so that, as of today, the enigma has still not been solved. Very little is known about this problem and, with all certainty, it will, as happened with Hilbert's 23 problems, become a benchmark of research in the 21st century. Because there is still a long way to go before the chaos of complex systems can be fully understood. And, as of today, God not only plays dice with the Universe, but also with graphs, numerical sequences and the new pathways that the future is laying for us.

Appendix

Proof of Sperner's lemma

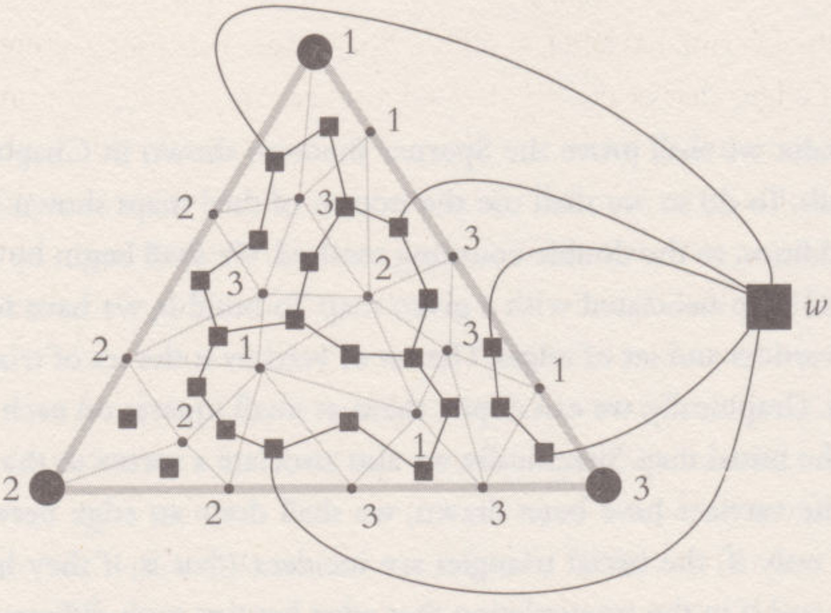
In this appendix we shall prove the Sperner theorem shown in Chapter 2 with all possible details. To do so we shall use the notion of dual maps shown in that same chapter, in addition to the double counting method. We shall begin by recalling the notion of dual map associated with a given map. To build it, we have to define this map's set of vertices and set of edges. The set of vertices is the set of triangles of our triangulation. Graphically, we can depict them as small squares on each of the small triangles of the initial map. Specifically, we also associate a vertex to the unbounded face. Once the vertices have been drawn, we shall draw an edge between two of them if, and only if, the initial triangles are incident (that is, if they have an edge in common) and if in the triangulation that edge has tips with different labels. This produces three different configurations – new vertices of degree 3, degree two and degree zero. Each of these cases is associated with the nature of the triangles in question, as is shown in the next figure:



Proposed dual construction for each type of triangle.

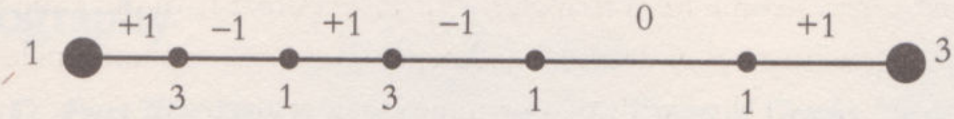
After building this new map, we can note that there are four types of vertices – isolated ones, those that are incident with two edges, those incident with three edges and the external vertex, which is special and to which we shall pay a little more attention.

Isolated vertices come from triangles with vertices that have the same label. In the same way, the vertices of degree 2 and degree 3 correspond to triangles whose vertices have been labelled with two or three labels respectively. These comments can be understood better by consulting the next figure:



Construction of the dual graph associated with the labelled triangulation, in accordance with the specifications obtained above.

Now let's pay a little more attention to the degree of the external vertex (we shall call it *w*) which, as we said above, is a little special. We are going to prove that the degree of this vertex is always odd (note in the previous figure that this vertex is degree 5). To do so, let's focus only on the vertices which are on the edges of the large triangle, and in particular on those that are on the edge determined by vertices 1 and 3. It should be remembered that those vertices, by hypothesis, can only be labelled by using two labels – 1 and 3. We can see the following: if we go to the vertex of the initial triangle whose label is 1, and we go towards the vertex of the initial triangle whose label is 3, then we will be crossing different vertices with labels 1 and 3. We associate a value +1 to a transition that goes from label 1 to label 3 (that is to say, to the corresponding edge) and, reciprocally, a value of -1 if the transition is made from label 3 to label 1. If there is no change of label, we associate a value of 0 to the corresponding edge. Let's look at the following example so as to clarify this point:



A concrete example of the construction when going from vertex 1 to vertex 3.

As the initial vertex has label 1 and the final vertex has label 3, there must be one more unit of edges that bear a +1 than edges that carry a −1. The number of edges with different tips (those that bear either a +1 or a −1) is an odd number, as it is the sum of two consecutive numbers. These edges are precisely those that contribute to the final degree of vertex w. Finally, by applying the same argument to the other two edges of the initial triangle, we reach the conclusion that the degree of w is the sum of three odd values and, consequently, its degree is odd.

Armed with these observations we now have all the ingredients for proving our result. To do so we shall use the double counting formula as applied on graphs, which told us that for any graph $G=(V,A)$ it holds that double the number of edges is equal to the sum of the degrees of its vertices. As we said before, the important thing about this formula is to note that the double of a number is always an even number. Therefore, the sum of the degrees of the vertices will also be an even number!

We shall now analyse the sum that corresponds to the sum of the degrees, and to do so let's look at each vertex's contribution to the total. Consequently, we start off from the previous sum in accordance with the classification that we did above:

$$d(w)+\sum_{v \text{ is degree } 0} d(v)+\sum_{v \text{ is degree } 2} d(v)+\sum_{v \text{ is degree } 3} d(v).$$

On the one hand, if a vertex is of degree 0, then $d(v)=0$, so in the previous sum we do not need to worry about the sum referring to the isolated vertices. To sum up, we have the following relationship:

$$d(w)+\sum_{v \text{ is degree } 2} d(v)+\sum_{v \text{ is degree } 3} d(v).$$

On the other hand, the sum associated with the vertices of degree 2 is always an even number, as each of the addends is a multiple of 2. To summarise, the sum

$$d(w) + \sum_{v \text{ is degree } 3} d(v)$$

must be an even number. As we saw above that $d(w)$ is an odd number, we can conclude that the sum

$$\sum_{v \text{ is degree } 3} d(v)$$

must also be an odd number and, consequently, not 0! Consequently, there exists another triangle with vertices labelled with different labels, which is precisely what we wanted to prove.

Bibliography

ALSINA, C., *From Tube Maps to Neural Networks: The Theory of Graphs,* Series: Everything is Mathematical, Barcelona, RBA, 2013.

ARNOLD, V., ATIYAH, M. et al. (editors), *Mathematics: Frontiers and Perspectives,* International Mathematical Union, 2000.

FEYNMAN, R., *Surely You're Joking, Mr. Feynman!,* New York, W.W. Norton and Company, 1997.

GOWERS, T. et al. (editors), *The Princeton Companion to Mathematics,* New Jersey-Oxford, Princeton University Press, 2008.

GRACIÁN, E., *Prime Numbers: An Unpredictable Series,* Series: Everything is Mathematical, Barcelona, RBA, 2013.

HOFFMAN, P., *The Man Who Loved Only Numbers. The Story of Paul Erdos and the Search for Mathematical Truth,* New York, Hyperion Books, 1998.

SCHECHTER, B., *My Brain Is Open. The Mathematical Journeys of Paul Erdos,* Oxford, Oxford University Press, 1998.

TUTTE, W., *Graph Theory as I Have Known It,* Oxford Lecture Series in Mathematics and its Applications, 1998.

Index

The Art of Counting

Enumeration and combinatorics

Many of the most important questions in modern mathematics require mastery of an extremely special art: counting. The branch of mathematics responsible for turning enumeration into an art form is called 'combinatorics', and thanks to legendary figures such as Paul Erdös, this field has formed the basis of some of the most amazing mathematical developments of the new millennium.